ISSN : 2456-3307

EAST POINT
GROUP OF INSTITUTIONS

Techno Science Academy
The International Open Access Publisher

IJSRCSEIT

National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology

22nd July 2023

East Point College of Engineering and Technology

Organized by

Department of Computer Science and Engineering,
East Point College of Engineering and Technology,
Bangalore, Karnataka, India

Volume 9, Issue 9, July-August-2023

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN COMPUTER SCIENCE, ENGINEERING AND INFORMATION TECHNOLOGY

Peer reviewed and Refereed International Scientific Research Journal

Impact Factor Value 7.254

Email: editor@ijsrcseit.com

National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology

On

22nd July 2023

EAST POINT GROUP OF INSTITUTIONS

Organized by

Department of Computer Science and Engineering, East Point College of Engineering and Technology, Bangalore, Karnataka, India

In association with

IJSRCSEIT

International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

Volume 9, Issue 9, July-August-2023

Published By

Techno Science Academy
The International Open Access Publisher

website : www.technoscienceacademy.com

CONTENTS

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

A Survey on Text Based Recommendation System

Asst. Prof. Pandey Nivedita*[1], Mr. Swapnil Mankar[2]
*[1]Department of CSE VTU University/East Point College Engineering and Technology, Bangalore, Karnataka, India
[2]R and D Department Bangalore, Karnataka, India

ABSTRACT
A wide range of textual material is produced by numerous websites on the Internet, including news, academic publications, ebooks, personal blogs, and user reviews. These websites include so much textual content that it can be difficult for users to find the information they need. To resolve this problem,The creation of "Text-based Recommendation Systems (RS)" is underway. They are the systems that can use text as their main feature to quickly find the pertinent information. There are numerous methods available for developing and assessing such systems. Although several surveys list the general characteristics of recommendation systems, a thorough literature assessment on text-based recommendation systems is still lacking. We examine the most recent research on text-based RS in this publication. The four primary facets of text-based recommendation systems employed in the reviewed literature are primarily covered by this survey. Datasets, feature extraction methods, computational frameworks, and evaluation measures make up the aspects. Publicly accessible datasets are thoroughly evaluated in this work since benchmark datasets are essential to any research. But in order to make these publicly available and proprietary datasets' properties more known to new researchers, we have combined them together. Additionally, the use of the feature extraction techniques from the text in the development of text-based RS is covered. The usage of these properties in various computational methods is later described. Some evaluation metrics are used to assess these systems. According to the report, Word Embedding is the most popular feature selection method currently being employed in research. We can also conclude that combining text features with other features improves the accuracy of the recommendations. The study emphasizes the fact that the majority of the effort focuses on English textual data and that the most common area is news recommendation.

Keywords—text based recommendation system. Filtering

I. INTRODUCTION

Recommender Systems (RSs) gather data on consumer preferences for a range of goods, including films, music, books, jokes, gadgets, software, websites, vacation spots, and e-learning materials. The data may be gathered directly (commonly through user ratings), or inadvertently (usually by surveillance of user behavior, such as music listened to, programmes downloaded, websites viewed, and books read). User demographics (such as gender, age, and nationality) may be used by RS. Web 2.0 regularly makes use of social media statistics

including follows, followed, twits, and postings. More and more people are using data from the Internet of Things, such as GPS coordinates, RFID, and live health signals.

RS employs a range of information sources to create forecasts and product recommendations for customers. When giving proposals, they try to strike a balance between elements like stability, dispersity, innovation, and correctness. Collaborative filtering (CF) techniques are important in the recommendation process, even if they are commonly used in conjunction with other filtering strategies like content-based, knowledge-based, or social ones. CF is founded on how people have made decisions throughout history: in addition to our own experiences, we also rely our choices on information and understanding that comes to each of us from a sizable network of acquaintances.

The collaborative, content-based, and demographic filtering types that were most popular at the start of the RS were discussed. Breese et al. assessed the CF predictive accuracy of several algorithms; thereafter, the classical publication defines the framework for Collaborative Filtering RS assessment.The importance of RS hybrid techniques, which combine many techniques to gain the benefits of each, has been demonstrated by the progress of the discipline. The hybrid RS has been the subject of a survey . The use of social-filtering, a method that has gained popularity recently through social networks, is not addressed, though.

At the start of the RS, the neighborhood-based CF has been the most widely used recommendation approach; Herlocker et al. present a set of criteria for creating neighborhood-based prediction systems. In their overview of the RS field, Adomavicius and Tuzhilin [3] highlight the most challenging areas where RS researchers should concentrate on the "next generation of RS": limited content analysis and overspecialization in content-based methods, cold-start and sparsity in CF methods, model-based techniques, non intrusiveness, flexibility (real-time customization), etc.

Our survey seeks to advance the evolution of the RS, moving from a first phase based on the traditional Web to the present second phase based on social Web, which is currently moving to a third phase (Internet of things), as opposed to the existing surveys, which concentrate on the most pertinent methods and algorithms of the RS field. In order to be helpful to the new readers of the RS area, we have incorporated certain conventional themes within this survey, including RS foundations, the k-Nearest Neighbours method, cold-start difficulties, similarity metrics, and RS evaluation. The remainder of the paper discusses brand-new subjects that are not included in current surveys. Advanced readers in RS will learn about social information (social filtering: followers, followed, trust, reputation, credibility, content-based filtering of social data; social tagging and taxonomies) in depth through this survey, as well as how to explain recommendations to groups of users. This survey will be helpful to readers who are interested in brand-new and upcoming applications because it provides information on the most recent works in location-aware RS trends and bio-inspired techniques. They will also learn about certain crucial topics, like teleoperation, telepresence, privacy, security, P2P information, and the utilization of the Internet of Things (RFID data, health parameters, surveillance data, etc.).

We provide a clear explanation of the process used to choose the most important papers in the RS field. The methods, algorithms, and models for recommending material based on information from the traditional web, including ratings, demographic information, and item data (CF, demographic filtering, content-based filtering, and hybrid filtering) are described. Measures for assessing the accuracy of the RS predictions and recommendations are described. The utilization of social information from Web 2.0 for recommendations through ideas like trust, reputation, and credibility is demonstrated. We will also discuss methods for obtaining social information based on content (such as tags and postings).

Online data has increased dramatically as a result of digital technological developments, particularly with the launch of smartphones. Social media platforms like Twitter and Facebook are important sources of data generation. Additionally, websites that answer questions, like Quora and Stack overflow, are rapidly adding data to this pool. Furthermore, during the past few years, both the trend of personal blogging and the quantity of publications have grown significantly. Users' lives have been impacted by the introduction of digitalization in both positive and negative ways. The data is readily and immediately accessible, which is a good thing.Recommendation systems have been created to address the issue of locating the most pertinent data among the overabundance of data [1].

II. SURVEY

It is a special collection of tools and methods that recommend to a user certain things that might be of interest to them. A recommendation system keeps track of a customer's profile and suggests a good or service based on their interests [2]. These recommendations can be found in any domain, from a news story to read to a web service to employ in software. The recommendation problem is divided into two parts, namely (i) calculating the value of prediction for each item and (ii) ranking these things according to their prediction value. And there are numerous methods to complete this work. They are most well-liked by (CF) and (CB) [1]. In CF, the user is presented with a new item via the recommendation system, which the user then consumes. In contrast, the recommendation system for CB suggests a new item based on its content and characteristics. In other words, the recommended item will share characteristics with the goods that the same person has already consumed. But recently, both of these methods have been combined and have produced encouraging outcomes. As any sort of data, including photos, text, numbers, and others, can be used, any type of data could be put into the recommendation system. The textual data makes up a sizeable share of the available data kinds. For instance, the main sources of this work include news, research publications, blogs, and various reports. The amount of textual data presents similar difficulties to those previously described for data in general.As a result, academics have also focused on textual data recommendation algorithms to discover the most pertinent content. Textual recommendations can be found in a variety of contexts, including news, articles, blogs, books, and movies. These text-based domains have been the focus of research for several decades [6]. There are unique problems in every domain [7]. As a result, each domain uses somewhat distinct methodologies. An overview of all such effort is therefore absolutely necessary. In practically every area of computer science, deep learning models have recently taken the role of conventional algorithms. Similarly, this new tendency has been embraced by the field of recommendation systems. For textual recommendations, there are several different neural network topologies that can be employed.

This assessment also provides a summary of the most recent methods used in textual recommendation systems. This page also compiles the evaluation methods for textual recommendation systems. Traditional assessment measures from the discipline of information retrieval are primarily used for RS. However, textual RS also makes use of some unusual measures like specificity and diversity [12].They are most well-liked by (CF) and (CB) [1]. In CF, a new item that is consumed by users who are similar to the user is recommended to them. In contrast, the recommendation system for CB suggests a new item based on its content and characteristics. In other words, the suggested item will share characteristics with the foods you've already eaten.

But recently, both of these methods have been combined and have produced encouraging outcomes. As any sort of data, including photos, text, numbers, and others, can be used, any type of data could be put into the recommendation system. The textual data makes up a sizeable share of the available data kinds. For instance, the main sources of this work include news, research publications, blogs, and various reports. The problems presented by the abundance of textual data are similar to those raised earlier for data in general. To locate the most pertinent material, academics have also focused on textual data recommendation algorithms. News recommendations, article/blog recommendations, book/movie recommendations, etc. are a few examples of textual recommendation domains. Work is being done by researchers on these text-based

III. OBJECTIVES FOR RESEARCH

The primary goals of this study are listed below.

- One of the main goals of the study is to provide the groundwork for future research on text-based recommendation systems, which has grown to be a significant research paradigm. The accomplishments related to textual data recommendations will be highlighted in the essay.
- To construct assessment metrics used in the evaluation of text-based recommendation systems;
- To gain understanding of how to extract meaningful features using various methodologies.
- To provide a general overview of the attributes present in the datasets utilized in textual RS. The purpose of this study is to provide a head start for researchers who wish to continue their work in the area of textual recommendation systems.

IV. EXISTING SYSTEM

There are numerous surveys that explore the various facets of recommendation systems. Some of them are all-encompassing, while others are focused on a particular topic, like in Chen et al. which lists all recommendation systems that rely on user reviews. A brief history of content-based recommendation systems was offered in another study that analyzed recent trends in content-based recommendation systems. The most recent trends were discussed in terms of data and algorithms. The poll said that Link Open Data (LOD), which is used to obtain additional meta-data attributes for an item, is now commonly used for data purposes.

In addition, more content is acquired through forums, user reviews, and tags; this content is classed as User Generated Data (UGD), and heterogeneous information networks are also explored as a source of content. The following methods were emphasized in terms of algorithms; Deep learning approaches, novel metadata encoding, word and document embeddings to find latent features in the text, and meta-path based methods are some of the methods used to represent a path that corresponds to a relationship between two things.

A different study concentrated exclusively on those RS that used ontologies in the creation of e-learning RS. All of the potential types of RS were briefly covered at the outset of the study. The writer summarizes all the papers at hand in terms of the ontologies employed, the ontology representation language, and suggests learning resources after giving a thorough explanation of ontologies and e-learning systems. A thorough analysis of the recommendation system that is solely based on user reviews (UR) or whose effectiveness has been enhanced by UR was conducted in another survey . The broad introduction to RS and its fundamental methodologies (Content-based, Rating-based Collaborative Filtering, and Preference-based Product Ranking)

were provided in the first section of the survey.
The second section of the poll covered the components of the reviews utilized in RS, including frequent terms, review themes, feature opinions, contextual opinions, comparative opinions, review emotions, and review helpfulness. All of the studies that used UR for user and product profiling for recommendations were thoroughly discussed in the following two sections of the survey, and in the final section, they discussed the practical implications of their findings for five dimensions: data quality, addition of new users, advancement of algorithms, profile-building, and product domain. One such recent study provides information on the use of text mining techniques in various recommendation systems.
Although the paper's primary focus is deep learning approaches and text-based RS is not specifically mentioned, it does cover all work on textual data that has been done using the DL methodology. A paper published by Batmaz et al. compiles deep learning methods utilized in RS. It gives a quick overview of the development methodologies but focuses more on the problems and difficulties that DL can help solve. Additionally, a structured form of the domains in which these models are used is also provided. There are also some studies that focus on particular domains, such as the one reported by [7].
Despite the fact that the text-based approaches discussed in this are quite brief and simple, it nonetheless provides an overview of the trend among those who work on News RS. The paper provides a summary of News RS algorithmic approaches, difficulties, and evaluation criteria. This work also contains condensed summaries of popular and openly accessible data sets. Another Study covers the difficulties and techniques in the news industry, although it is more general and seldom ever specific about any design on pure textual data.

V. METHODOLOGY

There are numerous surveys that explore the various facets of recommendation systems. Some of them are all-encompassing, while others are focused on a particular topic, like in Chen et al. , which lists all recommendation systems that rely on user reviews. A brief history of content-based recommendation systems was offered in another study that analyzed recent trends in content-based recommendation systems. The most recent trends were discussed in terms of data and algorithms. The poll said that Link Open Data (LOD), which is used to obtain additional meta-data attributes for an item, is now commonly used for data purposes.
All of the potential types of RS were briefly covered at the outset of the study. The writer summarizes all the papers at hand in terms of the ontologies employed, the ontology representation language, and suggested learning resources after giving a thorough explanation of ontologies and e-learning systems.A thorough analysis of the recommendation system that is solely based on user reviews (UR) or whose effectiveness has been enhanced by UR was conducted in another survey . The broad introduction to RS and its fundamental methodologies (Content-based, Rating-based Collaborative Filtering, and Preference-based Product Ranking) were provided in the first section of the survey. The second section of the poll covered the components of the reviews utilized in the study, including frequent terms, review themes, feature opinions, contextual opinions, comparative opinions, review emotions, and review helpfulness.
All of the studies that used UR for user and product profiling for recommendations were thoroughly discussed in the following two sections of the survey, and in the final section, they discussed the practical implications of their findings for five dimensions: data quality, addition of new users, advancement of algorithms, profile-building, and product domain. One such recent study provides information on the use of text mining

techniques in various recommendation systems. Deep learning has been more popular recently, even in the field of RS, and virtually everyone is using it to construct RS because of its successful and promising outcomes. All of these works were gathered in an overview on deep learning methods used in RS.

Although the paper's primary focus is deep learning approaches and text-based RS is not specifically mentioned, it does cover all work on textual data that has been done using the DL methodology. A paper published by Batmaz et al. compiles deep learning methods utilized in RS. It gives a quick overview of the development methodologies but focuses more on the problems and difficulties that DL can help solve. Additionally, a structured form of the domains in which these models are used is also provided.

VI. ALGORITHM

THE K NEAREST NEIGHBORS RECOMMENDATION ALGORITHM

The reference algorithm for the collaborative filtering recommendation process is the k Nearest Neighbours (kNN) algorithm. Its main benefits are simplicity and reassuringly reliable results; nonetheless, its biggest drawbacks are limited scaling and susceptibility to sparsity in RS databases. An overview of this algorithm function is given in this section.

The kNN-based CF is conceptually simple, has an easy implementation, and typically yields forecasts and suggestions of high caliber. Due to the high amount of sparsity in RS databases, however, similarity measures frequently run into processing issues (usually caused by a lack of mutual ratings when comparing users and items) and cold start scenarios (users and items with few rankings).

The kNN algorithm's poor scalability is yet another significant issue . The procedure of creating a neighborhood for an active user gets excessively slow as databases (like Netflix) grow larger; the similarity measure needs to be evaluated every time a new user registers in the database. The scalability issue is greatly reduced by the item to item variant of the kNN method [200]. In order to achieve this, neighbors are determined for each item; their top n similarity values are then recorded, and over time, predictions and suggestions are made using the stored data. Although ratings from earlier processing or storage are not included in the recorded information, obsolete information for products is less sensitive than for users.

The creation of measurements to precisely and accurately determine the present similarity for the users (or things) is a recurring theme in CF research. A number of statistical measures, including the Pearson correlation, cosine, constraint Pearson correlation, and mean squared differences, have historically been applied. Metrics have recently been developed to meet the limitations and quirks of RS. The relevance notion was created to give greater weight to users and products that were more pertinent. Additionally, a set of measures was created expressly to work well under cold-start conditions. Based on similarity metrics, the kNN algorithm operates. More information on the existing RS similarity measurements is provided in the following sections. According to each of these factors, the similarity approaches often calculate the similarity between two users, x and y (user to user) users' item ratings. The item to item kNN version computes the similarity between two items i and j.

A formal approach of the kNN algorithm may be found in . In this section, we will provide an illustrative example of this algorithm. The method for making recommendations is based on the following three steps:

(a) Using the selected similarity measure, we produce the set of k neighbors for the active user a. The k neighbors for a are the nearest k (similar) users to u.

(b) Once the set of k users (neighbors) similar to active a has been calculated, in order to obtain the prediction of item i on user a, one of the following aggregation approaches is often used: the average, the weighted sum and the adjusted weighted aggregation (deviation-from-mean).

(c) To obtain the top-n recommendations, we choose the n items, which provide most satisfaction to the active user according to our predictions.

RS have included social information (such as posts, blogs, tags, friends lists, followed and unfollowed individuals), which has become more prevalent as web 2.0 has grown. The RS is enhanced by the added contextual information. Because social information supports traditional memory-based information (user ratings), such as users connected by a network of trust exhibit significantly higher similarity on items and meta-data than non-connected users, the sparsity problem inherent in memory-based RS is improved. Researchers use social information for three main reasons: to produce or suggest new RS, to enhance the accuracy of forecasts and recommendations, and to clarify the most important connections between social information and collaborative processes.

A major subject of research in RS is trust and reputation. The social data now present in RS is directly tied to this topic. The following are the most popular methods for creating trust and reputation measurements: User trust (a): determining a user's trustworthiness based on the other users' explicit information or determining a user's trustworthiness based on implicit data gleaned from a social network.Users can introduce labels linked to objects in the social RS field. Folksonomies are collections of information spaces made up of the triples "user, item, tag." Folksonomies are fundamentally used in two ways:

(1) to build tag recommendation systems (RS based purely on tags);

(2) To employ tags to enhance recommendation processes.

Social Filtering

Through the identification of a community network or affinity network using the unique data that users provide (such as messages and web logs), social information can be obtained overtly or implicitly. It is possible to enhance the RS outcomes by employing solely user ratings, thereby fostering an implicit social network. To provide suggestions, one can blend implicit and explicit knowledge sources.

A trust-based CF can make advantage of the explicit social information to enhance the quality of suggestions. Different methods, such as trust propagation mechanisms, a "follow the leader" strategy, personality-based similarity measurements , trust networks distrust analysis, and dynamic trust based on the ant colonies metaphor can be used to produce or utilize trust information.

Through the identification of a community network or affinity network using the unique data that users provide (such as messages and web logs), social information can be obtained overtly or implicitly. It is possible to enhance the RS outcomes by employing solely user ratings, thereby fostering an implicit social network.

To provide suggestions, one can blend implicit and explicit knowledge sources.

A trust-based CF can make advantage of the explicit social information to enhance the quality of suggestions. Different methods, such as trust propagation mechanisms , a "follow the leader" strategy, personality-based similarity measurements , trust networks distrust analysis, and dynamic trust based on the ant colonies metaphor , can be used to produce or utilize trust information.

They therefore propose that using the user's resemblance and familiarity with the people who have rated the things can help with judgment and decision-making. By combining social network data into CF, Fengkun and

Hong created a method to improve the effectiveness of recommendations. They obtained information about users' social network connections and preference ratings from a social networking website. After that, they assessed how well CF performed with various neighbor groups, which combined friends and the closest neighbors. According to Carmagnola et al. , joining a network exposes people to social dynamics that can affect their attitudes, behaviors, and preferences: SoNARS, a recommendation algorithm for content in social RS, is presented. SoNARS seeks out users who are active on social networks.

A recommendation system is what?

Numerous applications exist where websites gather user data and use that data to forecast the preferences of their users. They can then recommend the content they enjoy. Recommender systems are a technique of recommending things and concepts that are comparable to a user's particular way of thinking.

VII. CONTENT-BASED FILTERING

The goal of content-based filtering (CBF) is to suggest to the active user goods that have previously received favorable ratings. It is predicated on the idea that objects with comparable features will receive comparable ratings and .For instance, if a user enjoys a website that has the phrases "car," "engine," and "gasoline," the CBF will suggest further websites on the automotive industry.

As RS incorporates data on objects from users operating in web 2.0 contexts, such as tags, posts, opinions, and multimedia content, CBF is becoming increasingly significant.Overspecialization and insufficient content analysis are two difficult issues for content-based filtering [3]. The first issue is that it is challenging to automatically extract trustworthy data from a variety of resources (such as photographs, video, audio, and text), which can significantly lower the quality of suggestions. The second issue, known as overspecialization, is when users only receive recommendations for products that are extremely similar to the products they liked or preferred. As a result, users miss out on suggestions for products they might like but have never heard of (for example, when a user only receives recommendations for fiction films). The uniqueness of recommendations can be assessed .

The qualities of the products you want to recommend must be retrieved for CBF to function . Depending on the item's domain, a set of attributes is often explicitly established for each one. Classic information retrieval techniques must be employed to automatically construct such attributes in some situations, such as when it is wanted to recommend textual material (e.g., term frequency, inverse document frequency, and normalization to page length).

The CBF mechanism is depicted in Fig. and consists of the following steps: To make recommendations, one must first extract the properties of the products, then compare those features to the preferences of the active user, and finally suggest goods having those characteristics.

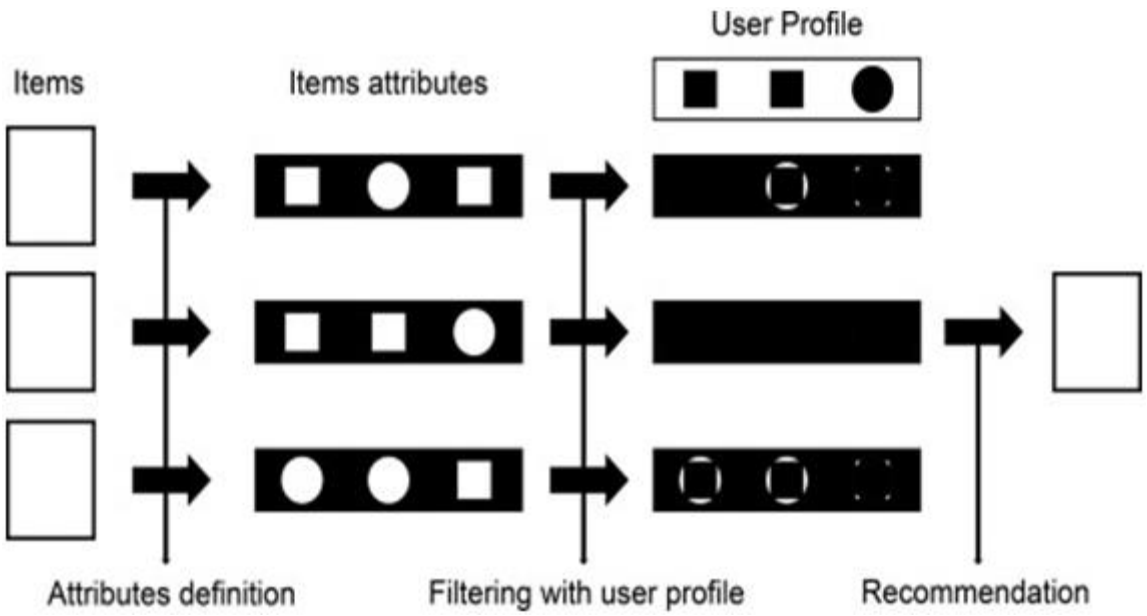


Figure no:1

The main goal of CBF is to ascertain if a user will like a particular thing once the attributes of the objects and user profiles are understood. Traditionally, heuristic techniques or classification algorithms, such as rule induction or nearest neighbors methods, are used to handle this problem. linear classifiers and probabilistic techniques are used in Rocchio's approach.The pure CBF has a number of drawbacks:

A) Generating the attributes for things in some domains is a challenging task.
B) Because CBF tends to promote the same kinds of products by default, it has an overspecialization problem.
C) It is more challenging to get user feedback because, unlike CF, users of CBF don't frequently review the things, making it impossible to know whether the recommendation is accurate.

VIII. COLLABORATIVE BASED FILTERING

In Collaborative Filtering, we tend to find similar users and recommend what similar users like. In this type of recommendation system, we don’t use the features of the item to recommend it, rather we classify the users into clusters of similar types and recommend each user according to the preference of its cluster.

There are basically four types of algorithms o say techniques to build Collaborative filtering based recommender systems:

- Memory-Based
- Model-Based
- Hybrid
- Deep Learning

IX. CONCLUSION

Text-based recommendation systems have become more prevalent in the last decade because the internet is generating the bulk of textual data every day over different websites. This study aims to explore text-based

recommendation literature and summarize critical approaches to provide a single platform for the understanding of new researchers. The survey covers four main aspects of a text-based recommendation system. First, what are the fundamental techniques of feature extraction used in text-based recommendation systems.Second, proprietary and publicly available datasets and their details. Third, how such systems are evaluated, what are the most frequently used evaluation metrics, and lastly what algorithmic approaches are opted to formulate the problem. Second, proprietary and publicly available datasets and their details. Third, how such systems are evaluated, what are the most frequently used evaluation metrics, and lastly what algorithmic approaches are opted to formulate the problem.

X. REFERENCES

[1]. J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," Knowl.-Based Syst., vol. 46, pp. 109–132, Jul. 2013.

[2]. Y. Wang, N. Stash, L. Aroyo, L. Hollink, and G. Schreiber, "Semantic relations for content-based recommendations," in Proc. 5th Int. Conf. Knowl. Capture (K-CAP), 2009, pp. 209–210.

[3]. Z. Cao, X. Qiao, S. Jiang, and X. Zhang, "An efficient knowledge-graphbased Web service recommendation algorithm," Symmetry, vol. 11, no. 3, p. 392, Mar. 2019.

[4]. X. Kong, M. Mao, W. Wang, J. Liu, and B. Xu, "VOPRec: Vector representation learning of papers with text information and structural identity for recommendation," IEEE Trans. Emerg. Topics Comput., early access, Apr. 26, 2019, doi: 10.1109/TETC.2018.2830698.

[5]. V. Setty and K. Hose, "Event2Vec: Neural embeddings for news events," in Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., New York, NY, USA, 2018, pp. 1013–1016. [6] J. Beel, S. Langer, B. Gipp, and A. Nürnberger, "The architecture and datasets of docear's research paper recommender system," D-Lib Mag., vol. 20, no. 11/12, 2014.

[6]. J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems: A literature survey," Int. J. Digit. Libraries, vol. 17, no. 4, pp. 305–338, Nov. 2016.

[7]. C. Musto, T. Franza, G. Semeraro, M. de Gemmis, and P. Lops, "Deep content-based recommender systems exploiting recurrent neural networks and linked open data," in Proc. Adjunct Publication 26th Conf. User Modeling, Adaptation Personalization, New York, NY, USA, 2018, pp. 239–244.

[8]. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," J. Mach. Learn. Res., vol. 11, no. 12, pp. 3371–3408, 2010.

[9]. G. de Souza Pereira Moreira, F. Ferreira, and A. M. da Cunha, "News session-based recommendations using deep neural networks," in Proc. 3rd Workshop Deep Learn. Rec. Syst., New York, NY, USA, 2018, pp. 15–23.

[10]. G. Shani and A. Gunawardana, Evaluating Recommendation Systems. Boston, MA, USA: Springer, 2011, pp. 257–297. [12] D. Jannach, L. Lerche, I. Kamehkhosh, and M. Jugovac, "What recommenders recommend: An analysis of recommendation biases and possible countermeasures," User Model. User-Adapted Interact., vol. 25, no. 5, pp. 427–491, Dec. 2015.

[11]. L. Chen, G. Chen, and F. Wang, "Recommender systems based on user reviews: The state of the art," User Model. User-Adapted Interact., vol. 25, no. 2, pp. 99–154, Jun. 2015

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Data Mining In Higher Education- Persistence Clustering and Prediction

Dr. R. Senkamalavalli[1], Shammi.L[2]
[1]Associate Professor, [2]Assistant Professor
Department of CSE, East Point College of Engineering and Technology, Bangalore, Karnataka, India

ABSTRACT

Data mining is slowly but surely making its way into the educational field after dominating the business fields. Higher education will find larger and wider applications for data mining than its counterpart in the business sector, because higher education institutions carry
their duties that are data mining
Intensive: scientific research that relates to the creation of knowledge, teaching those concerns with the transmission of knowledge, and institutional research that pertains to the use of knowledge for decision making. All the above tasks are well within the boundaries of knowledge management, which drives the need for better and faster decision-making tools and methods.
Data mining enable organizations to use their current reporting capabilities to uncover and understand hidden patterns in vast databases. These patterns are then built into data mining models and used to predict individual behavior with high accuracy.
Data mining uses a combination of an explicit knowledge base, sophisticated analytical skills, and domain knowledge to uncover hidden trends and patterns. These trends and patterns form the basis of predictive models that enable analysts to produce new observations from existing data.
Keywords: Decision Making, Higher Education, predictive Model.

I. INTRODUCTION

One of the biggest challenges that is faced by higher education today is predicting the paths of students and alumni institutions would like to know, for example, which students will enroll in particular course programs, and which students will need assistance in order to get placements etc.? In addition to this challenge, traditional issues such as management, financial assistance and time- to-degree continue to motivate higher education institutions to search for better solutions.
One way to effectively address these students and alumni challenges is through the analysis and presentation of data, or data mining.

II. RELATED WORK

In 2007, Vranice et al., in [2] explored data mining algorithms on a Croatian university students' data. The focus was whether future students of this course will succeed or fail. They have tested several algorithms and their result was similar, however, the authors indicated that their sample was small and perhaps future research would include more detailed student data.

Parack et al., in [3] presented a paper focusing on predicting academic trends and student patterns behavior. This has eased the process in grouping similar student profiles and identifying their learning patterns.

Very interesting paper by Nasiri and Minaei in [4] in higher education focused on two issues: GPA and academic dismissal in a Learning Management System (LMS). Both algorithms used for the data mining process indicated a weakness if there is a slight variation to the data, this will lead to different results, they solved it by adding association rules.

Shi et al., presented a paper [5] focusing on managing the university curricula based on data association mining technology. They reached the conclusion that if a student was successful in a certain course, then he will be Successful in similar courses as well. An example can be mathematics and physics.

Again in India, Bunkar et al., presented a paper where they applied data mining techniques to predict the performance improvement of graduate students using classification [6]. Several techniques were discussed and the authors were able to isolate students that are most likely to fail and provide proper counseling and guidance.

In Romania, Bresfelean et al., exploited the university academic failure issue in [7]. Their aim was to define an academic failure profile for students to be able to predict students' exam failure and success based on data mining techniques. They aimed to improve students learning methods and detect their weakness, and assist in managerial educational decision.

Offering high quality education means being able to predict student enrollment in courses, identify beneficial teaching methods, forecast student performance in end exams and identify drop out rates, and help those students during the semester. Their method described in [8] based on classification helped in the proper dividing of students, and paying special attention to students most likely to fail, and help in increasing the success in the success and failure ration.

In China, Wu had a different intake on higher education by using clustering to identify student course selection based on teachers [9]. Their goal was based on guiding the students and giving them most appropriate advice to succeed. The author identifies different categories of student-teacher selection, out of the three clusters, one was successful in selecting the teacher based on several important factors which has increased their class interactivity, discipline, behavior and led to their success. The other two groups, should receive proper guidance to achieve what the first group has achieved.

Japan had another take on the topic by focusing on the university curricula and built an EDM to reach an optimal learning success in terms of best possible Grade Point Average (GPA) [10]. They enforced their system by including an individual learner profile that includes pre- university educational data and grows as the student progresses through the university. This helps in grouping similar profiles and inferring success patterns that can't be identified through conventional student analyzing methods.

A recent study in 2012 in [11] focused on predicting the drop-out rate of students from universities, colleges, and institutions in developing countries. Mustafa et al., used classification and regression to classify successful

and unsuccessful students based on gender, financial condition, ethnicity, work-status, disability, and study environment. The search was based on background information. And they identified the most important classification factors were: financial support, age group, and gender.

In 2005, a paper presented by Delavari et al., developed an educational data mining model to be tested in a Malaysian University [12]. They had three targets, the first was understanding the course enrollment pattern in a course, and identifying which students are successful in passing the course, and who will fail. This leads to their next target, faculty, who take proper action in guiding these students, and choose direct or indirect methods to provide the necessary class skills, and financial aid. The final target of the system, are the decision makers, the system enhances the education quality and provide quality management, improve policy making, and setting new strategies and goals.

More on the topic of student failure in college courses can be found in [13]. The author used association rules mining algorithm to find the factors that lead to student course failures.

In Spain, a research presented by Tovar and Soto to improve their predicting model [14]. Instructors can locate students having problems with the course and help them. It also identifies student who have the capabilities to pass the course but fail. Their research is based more on statistics than data mining techniques.

A paper by Knauf et al., [15] presented their storyboard model which students at the Tokyo Denki University were using for progressing through the curriculum. Then they used data mining techniques to build an individual student profile to record their personal properties, talents, weaknesses, and preferences. This model presents students with suggested courses where they will be successful.

Ningning presented a paper focusing on data warehousing and data mining [16]. His model identifies students most likely to drop out a course. This can aid business managers in pinpointing students in need of help and guidance.

III. KNOWLEDGE MANAGEMENT DRIVING DATA MINING

Data mining should be performed on very large or raw datasets using either supervised or unsupervised data mining algorithms.

Several authors have written about the factors behind the down of data mining. For instance, Sir.Therling identified three reasons viz;

- The ease of data collection and storage
- The computing power of modern processors and
- The need for fast and real time data mining.

Yet one important reason absent from these is the growing interest in knowledge management. Knowledge, a focal point of ontology or epistemology, is the product of moving from data to information and finally to knowledge.

IV. DATA MINING MODELS

The following model, tiered knowledge management model (TKMM) developed by Jing Luan illustrates the dichotomous nature of modern knowledge management framework for higher education research professionals.

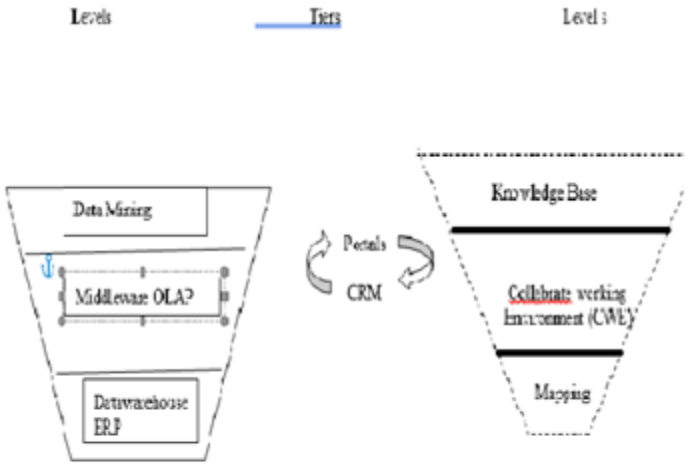


Figure: 4.1Tiered Knowledge Management Model (TKMM)

Models house the steps modules, and resources of the data mining process. Some data mining models include the entire process for a particular purpose, be it to cluster or predict.

Data mining is a powerful tool for academic intervention. The components of knowledge management are explicit (documented, measurable) and tacit (subjective, qualitative).

Documented, measurable explicit knowledge is most familiar and available to us, as it exists mostly in databases and other similar medium. While tacit knowledge, an entity of feelings, personalities and aptitudes is crucially important but it is hard to quantify and we will leave that for further study. All these components in customer relationship management (CRM) operational, analytical and collaborative are key users of data mining.

On the explicit side, data mining reflects the highest level of knowledge attainment that requires skills in data domain (tier one). Data querying and presentation (tier two). And artificial intelligence /machine learning (tier three). Data mining occupies the top tier and is dependent on the lower tiers. The following chart is a topology of the explicit knowledge of TKMM that illustrative in detail the relationships among three tiers and the software programs for each:

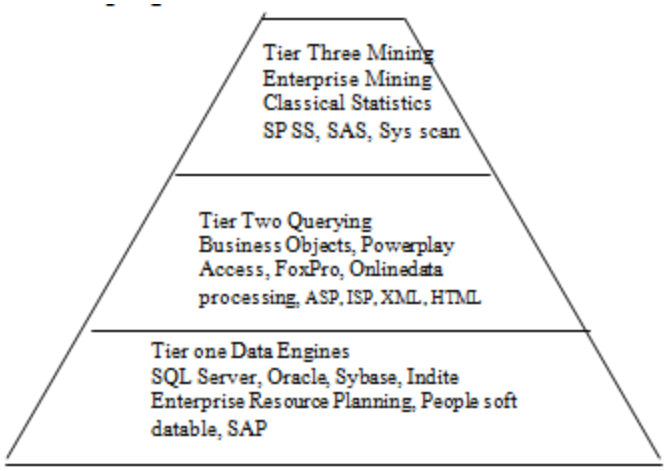


Figure 4.2: Topography of Tiered Knowledge Management Model (TKMM) for explicit Knowledge

4.1.1 SUPERVISED AND UNSUPERVISED MODELING

Classification and estimation use either unsupervised or supervised modeling techniques. Unsupervised data mining is used for situations in which particular groupings or patterns are unknown. In student course databases, for example, little is known about which courses are usually taken as a group, or which courses types are associated with which student types. Unsupervised data mining is often used first to study patterns and search for previously hidden patterns, in order to understand, classify, typify and code the objects of study before applying theories. Supervised data mining, however, is used with records that have a known outcome. A graduation database, for example, contains records of students who completed their studies, as well as of those who dropped out. Supervised data mining is used to study the academic behavior of both groups, with the intention of linking behavior patterns to academic histories and other recorded information.

V. DATA MINING IN HIGHER EDUCATION

Data mining is already fundamental to the private sector. Many of the data mining techniques used in the corporate world, however, are transferable to higher education.

What are the transferable techniques in data mining that are readily applicable in higher education? Infact, there are many algorithms are similar in concept to stored producers in object related programming in that they are universally applicable.

The following case study illustrates a key application of data mining in higher education.

Case Study: **Creating meaningful learning outcome typologies**

"What do institutions know about their students?" If the answer is a recital of enrolment percentages of the basic counts, institutions do not know their students as well as they could. This case study demonstrates how suburban community colleges can establish learning outcome typologies for students using unsupervised data mining.

SOLUTION

To establish appropriate typologies, we can use two step and k-means clustering algorithms we first will apply the algorithms to the general groupings that identifies students as "transfer oriented", "basic skill upgraded", " on campus directed", "vocational education directed" postal course oriented " etc, Note that the boundaries among clusters were under and dispersed.

Now we need to have a replacement method that takes care of cases, which do not appear to or belong to any group. Defining educational outcomes is easier said that done. Dropping out is also an outcome by itself. Further we need to determine the length of study which required decisions on how to deal with 'stop outs' students who left colleges and later returned. All of these situations test the data miners domain knowledge.

The two-step algorithm will produce following clusters: "Transfers", "Vocational students", "On campus directed", "Basic skills students", "students with mixed outcomes", and "Drop outs". K-means will validate this cluster. Some transfer students can complete their studies quickly some vocational can take it longer falls within this.

VI. RESULTS

If data mining can quickly identify potential donors by a ratio of two to four (correctly predicting two out of four who will donate) then the university can achieve results
By mailing only to the indicated 40% of the alumni donor population thus saving considerable time and money.

VII. CONCLUSION

With the ability to uncover hidden patterns in large databases, colleges and universities can build models that predict with a high degree of accuracy the behavior or several clusters. By acting on these predictive models educational institutions can effectively address issues ranging from transfers and retention to marketing and alumni relations. Data mining is a new type of exploratory and predictive data analysis that has tremendous applications in higher education institutional research alone. An effective data mining cycle was presented to clarify the process. We will take this research further by incorporating RULES-3 developed by Pham and Aksoy in [17] and later improved by Mathkour in [18], in the Edu Gate System and develop a tool for management to make their insightful decisions. Moreover, we will also focus on data selection, and preparation phases when the system is tested for accuracy, validation, and verification.

VIII. REFERENCES

[1]. D. A. Alhammadi and M. S. Aksoy, "Data Mining in Education – An Experimental Study," International Journal of Computer Applications, vol. 62, no. 15, pp. 31-34, 2013.
[2]. M. Vranic, D. Pintar and Z. Skocir, "The use of data mining in education environment," IEEE Telecommunications ConTel 2007. 9th International Conference, pp. 243-250, 2007.
[3]. S. Parack, Z. Zahid and F. Merchant, "Applications of Data Mining in Educational Databases for Predicting Academic Trends and Patterns," IEEE International Conference on Technology Enhanced Education (ICTEE), pp. 1-4, 2012.
[4]. M. Nasiri and B. Minaei, "Predicting GPA and Academic Dismissal in LMS sing Educational Data Mining: A Case Mining, " IEEE Third International Conference on E- Learning and E-Teaching (ICELET), pp. 53-58, 2012.
[5]. F. Shi, Q. Miao and D. Mei, "The Application of Data Association Mining Technology in University Curriculum Management," IEEE Symposium on Robotics and Applications (ISRA), pp. 521-524, 2012.
[6]. K. Bunkar, U. K. Singh, B. Pandya and R. Bunkar, "Data Mining: Prediction for Performance Improvement of Graduate Students using Classification," IEEE Wireless and Optical Communications Networks (WOCN), pp. 1-5, 2012.
[7]. V. P. Bresfelean, M. Bresfelean, N. Ghisoiu and C.A. Comes, "Determining Students' Academic Failure Profile Founded on Data Mining Methods," IEEE 30th International Conference on Information Technology Interfaces, pp. 317-322, 2008.
[8]. S. K. Yadav, B.K. Bharadwaj and S. Pal, "Data Mining Applications: A comparative study for Predicting Student's Performance", International Journal of Innovative Technology and Creative Engineering (IJITCE), vol. 1, no. 12, pp. 13-19, 2011.

[9]. F. Wu, "Apply Data Mining to Students' Choosing Teachers Under Complete Credit Hour," IEEE Education Technology and Computer Science (ETCS), vol. 1, pp. 606-609, 2010.

[10]. M. N. Mustafa, L. Chowdhury and M. Kamal, "Students Dropout Prediction for Intelligent System from Tertiary Level in Developing Country," International Conference Informatics, Electronics & Vision (ICIEV), pp. 113-118, 2012.

[11]. N. Delavari, M. R. Beikzadeh and S. Phon-Amnuaisuk, "Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System," IEEE International Conference on Information Technology Based Higher Education and Training (ITHET), pp. F4B/1-F4B/6, 2005.

[12]. Z. Zhang, "Study and Analysis of Data Mining Technology in College Courses Students Failed," International Conference on Intelligent Computing and Integrated Systems (ICISS), pp. 800-802, 2010.

[13]. E. Tovar and O. Soto, "The Use of Competences Assessment to Predict the Performance of First Year Students," IEEE Frontiers in Education Conference (FIE), pp. F3J-1, 2010.

[14]. R. Knauf, Y. Sakurai, K. Takada and S. Tsuruta, "Personalizing Learning Processes by Data Mining," IEEE International Conference on Advanced Learning Technologies (ICALT), pp. 488-492, 2010.

[15]. G. Ningning, "Proposing Data Warehouse and Data Mining in Teaching Management Research," International Forum on Information Technology and Applications (IFITA), vol. 1, pp. 436-439, 2010.

[16]. D. T. Pham and M. S.Aksoy, "RULES: A Simple Rule Extraction System," Expert Systems with Applications, vol. 8, no. 1, pp. 59-65, 1995.

[17]. H. Mathkour, "RULES3-EXT: Improvements of RULES3 Induction Algorithm," Mathematical and Computational Applications, vol. 15, no. 3, pp. 318-324, 2010.

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Blockchain Based Decentralized Social-media to Prevent False Copyright Infringement

Divya U H[1], Kavitha[2], Manjula[2]
[1]Assistant Professor, [2]Student
Department of CSE, EPCET, Bangalore, Karnataka, India

ABSTRACT

Social Media Networks have become a vital part of everyone's daily life. Social Media plays a crucial role in connecting people. Most of the social media sites like YouTube, WhatsApp, Twitter, etc. that exist today operate in a centralised manner. It means a single organisation is responsible for managing all the activities, like the regulation of business rules, terms and conditions, maintenance, and development of social media. These companies solely control the content removal with respect to copyright and other false content. We provide a Decentralised system where people who use the platform—social media like SteemIt, Ushare, Mastodon, etc.—get power to manage the critical aspects like features needed for the platform's development and the reduction of false copyright infringement.

Keywords: Decentralized social media, Blockchain

I. INTRODUCTION

In recent days, most people depend on social media for a living. For example, professional YouTubers, Instagrammers, etc. These platforms handle copyright infringement in a centralised manner, which means if a user finds his or her content posted by another user, he files a copyright report, which will be evaluated by the company. If the company finds the report to be legitimate, it will discard the copyrighted post, and the user who posted the post will be demonetized or penalised in other ways. Dapps uses PoS, DPoS, PoT, etc. for consensus. It has created a major problem as the copyright content might be fake and a single central company has no time to evaluate all the reports. Modern social media sites like YouTube assume that a user with a higher number of subscribers is legitimate. This has led to many legitimate YouTubers getting demonetized, which can affect their major source of income. On a centralised platform, companies with a single brand make the mark of the company's popularity and reputation, which holds the trust of the people on the platform. They control the activities, such as launching a new feature or event that can officially represent the platform and encourage people to use it more often. Centralization gives control to the company or stakeholders to apply their ideas, which they believe are best for the people.

To combat this problem of copyright infringement, we are designing a decentralised social media platform based on the principles of blockchain. Social media will not be controlled by a single person or organisation.

Control is distributed into the hands of several witnesses who are elected in an election similar to DPOS. Everyone who is part of the process will get a proportion of the reward based on their nature of work.

II. LITERATURE SURVEY

A. Incentivized Blockchain-based Social Media Platforms: A Case Study of Steemit

Steemit is operated by a decentralized community, where 21 members are periodically elected to cooperatively operate the platform through the Delegated Proof-of-Stake (DPoS) consensus protocol proposed by Li et al. [1]. It includes the decentralisation of data generated on the platforms and the deep integration of social platforms with the underlying cryptocurrency transfer networks. It is also the first blockchain-powered social media platform that incentivizes both creators of user-generated content and content curators. A few of the advantages of Steemit are that the data stored on the blockchain is publicly accessible and hard to manipulate. Also, on Steemit, users can maintain complete anonymity. Steemit uses its native cryptocurrency, STEEM. Thus, it is not dependent on any other platforms. It is completely censorship-free. But this could be misused also; hence, it has a few limitations. If the size of the group is too large, it won't be feasible foronly21memberstovalidateallposts.Powerofbigshareholders: It is observed that most of the time, the 21membersareconstitutedbybigstakeholders,whichsuppressesdecentralisation.RewardsystemsinSteemitmay also be misused by some users in ways that deviate from theoriginal intended goal of Steemit, such as buying votes frombots to promote some meaningless posts for profit.

B. Social-Chain: Decentralized Trust Evaluation Based on Blockchain in Pervasive Social Networking

The authors of [2] proposed the Proof-of-Trust consensus mechanism, which is lightweight and thus can be feasibly deployed in a mass of resource-limited PSN nodes. They proved the security of the Social Chain, which overcomes the risk of centralization and fork issues appearing in many existing blockchain systems. The experimental results further show its effectiveness and efficiency. Mining Winner Selection uniquely selects a block from multiple candidates, so a blockchain fork can be avoided. Specifically, they limit the total number of wins by an individual miner in a specific period to ensure decentralisation. Peng et al. [2] proposed that PoT, a newly generated block, can be confirmed as the next block if and only if it is approved by a sufficient number of miners with a sufficient sum of trust values. Miners can determine the correctness of the blockchain by verifying hash values. In PSN, it lacks a centralised party to perform information collection, social data aggregation, and trust evaluation, which should be self-organised by involved parties in practise. Proof-of-Trust can be easily hijacked if the number of malicious users is greater than a certain threshold. Most of the existing consensus mechanisms may not be applicable as they use cryptocurrency as an incentive. They cannot solve the problems of centralization and forking at the same time.

C. A Blockchain Enhanced Framework for Social Networking

User activities in SNSs are stored in the blockchain, along with queries for the data that are generated by APIs. Murimi et al. [3] implemented a blockchain that stores information about user content, preferences for sharing, rewards for sharing content, and records about data access. They define transactions as the set of actions performed on various websites. The user can choose the sharing preferences for her data. Thus, sharing preferences are not just dictated by the network or website settings on privacy and sharing. A user can control

the subset of her friends on a SNS that can access her data and can choose what portion of her anonymous data is available for access by other users. This framework is capable of data attribution on both anonymous and non-anonymous networks. It can be used to effectively track the number of users that engage with content of various kinds. Users can be rewarded for their transactions on the network by choosing reward algorithms that are suited to their privacy and monetization preferences. These rewards can be in the form of digital tokens and are also stored on the blockchain in the BEV-SNS version. Even though the user wants to stay anonymous, he will expose quite a lot of personal data in his posts. Anonymity and monetization cannot be achieved at the same time. It still has the problem of resource consumption for consensus as it uses Proof-of-Work.

D. Ushare: user controlled social media based on blockchain

In this system, users would be able to share their data with their circle of friends, family, and others. Chakravorty et al. [4] introduced a Personal Certificate Authority (PCA) for each user that would remain outside the blockchain in their personal space as client software. The PCA would issue certificates based on the circles created by a user to share their data. This allows only members belonging to a user's particular circle to view the content shared with that circle on the blockchain. Ushare consists of four key components: the blockchain, a hash table with encrypted content shared by a user, a Turing complete relationship system to control the maximum number of shares performed by the user's circle members, and a local PCA that manages the user's circles and encrypts data to be shared. The PCA creates an encrypted version of the data with the circle's public key and stores it in a distributed hash table. The user shares the hash ID of the encrypted data with each member of their circle. This enables the maintenance of precise traceability and control over shareability. The distributed hash table stores posts hashed at both the user level and group level, making them more secure and available only to members of that group. With the growth of the blockchain with multiple circles and members, key management issues could have a major impact on security and performance. It still has problems with the consensus algorithm, which was discussed in previous papers.

III. THE PROBLEM OF FALSE COPY RIGHT INFRINGEMENT

A. Copyright Management in Centralized social media

In modern-day centralized social media, a single organization is involved in taking down copyrighted content. As a result, a single organization has to perform all the validation of the reported content, which becomes the bottleneck in the system. So, centralized social media has chosen an easy way of dealing with this problem, which is to favor the user or content creator who has more subscribers or followers on their platform. The problem with this approach is that if a user with large subscribers decides to claim others' original work, social media will favor the user with large subscribers instead of favoring the person with original work. This has become an everyday problem at the moment. So, a different approach has to be taken to eliminate this problem.

B. Decentralizing the socialmedia

Decentralising can be applied at various levels in a system. Some of the possibilities are

- Physical decentralisation (P2P networks): It involves decentralisation at the server. Any group of people can host their own server and start using social media. They can choose to remain anonymous or have open communication with people on other servers. e.g.: Mastodon.

- Logical decentralisation: Unlike physical decentralisation, there will be a single unit of servers connecting all the users around the world, but the decentralisation is applied at the data storage level. It could be storing the entire data of the network in a blockchain-based database, as shown in Figure 1, instead of a traditional database, or using the principles of blockchain to monitor the problems. e.g.: Steemit.

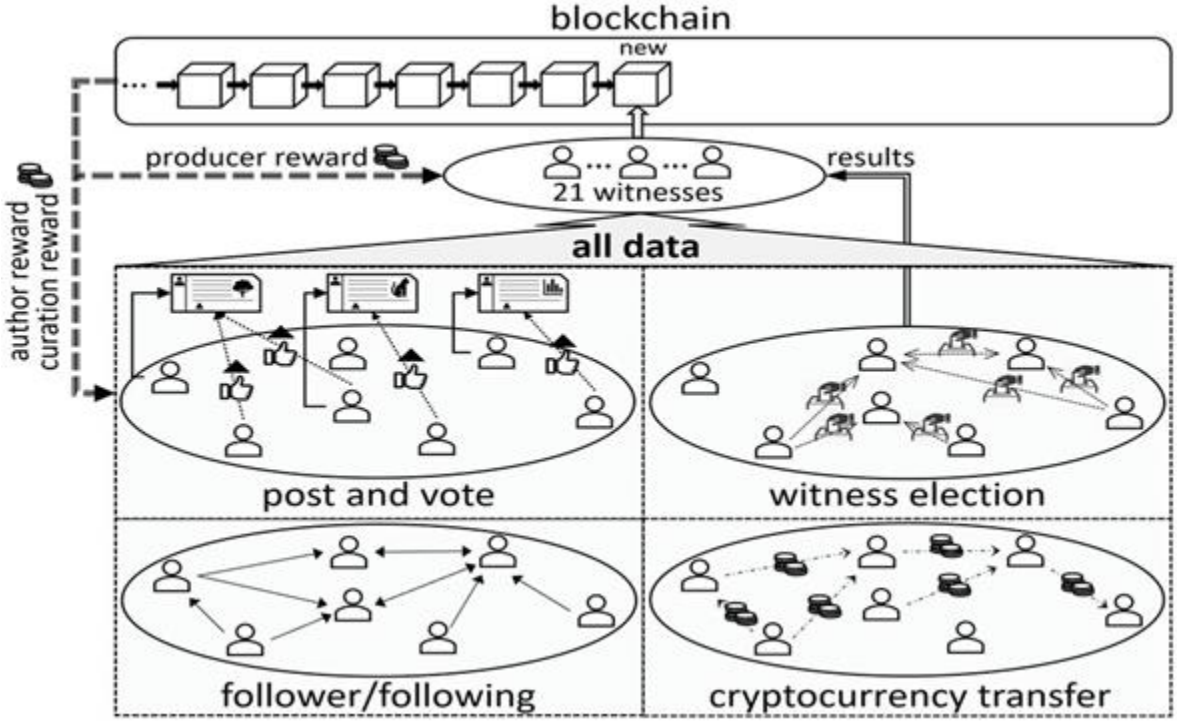


Figure1:BlockchainLogicalDecentralization

IV. DATASTORAGE

It is known that a lot of data is generated on social media. Based on the nature of the data, we may need to store it in different ways. There are posts, comments, user data, user activities, and cryptocurrency exchanges; we will suitably categorize and store them as per the needs. Blockchain design becomes a crucial part of our social media. There are various design approaches, so it is important to decide the tradeoffs of these designs. The proposed model uses blockchain as an immutable ledger. Instead of storing every social media activity on blockchain, which will unnecessarily increase the storage space. We are storing only those activities that are required to be stored as immutable. Some of them include witnesses validating posts, election data, cryptocurrency transfers, etc.

A. Abbreviations and Acronyms

- Dapps: DecentralizedApps
- PoT: ProofofTrust
- PoS: ProofofStake
- DPOS: DelegatedProofofStake
- SNS: SocialNetworkingService

B. Equations

A user in the platform can have the power vested in it and gain rewards due to up votes to the post. Up vote and Down vote will consider the user's powers to gain rewards.

Cost of 1 vote = (User's Power / Total Power in thePlatform)*k

k-Numbertocalculatethecostof1votewhichisgreaterthan1

Users get power p for each upvote by a person upv in aplatformconsideringthepreviousupvotesupvpreviousandtheirpower.

p=A*p+upvprevious

A-MovingAverageConstant

The total revenue generated is distributed in two ways. 60% isdistributedtostakeholdersandplatformmaintainers,and 40%isheld as stake convertible power P. Power can be convertedintomoneyK(Moneyperunitpower).

- Money M obtained through the conversion of a certain amountofpowerp,M=p*K
- Powerleftp'afterconversionintomoney,p'=P-p

The top20%ofthepeopleontheplatformcanparticipateintheelection. 5% of them will be elected as miners up to amaximum of 25 members. The number of miners is always odd.In the event of a copyright infringement issue occurrence,minerswillvotetodecidetherighteousowner.Theonewiththe maximum number of votes will be considered therighteousownerofthecontent.

V. USER GROUPS

There will be mainly two user classes: miners and normal users. Normal users can post content and upvote or downvote others' posts. They can join any group of interest, and their goal is to earn more reward points and become a miner in their group. Miners have the additional responsibility of validating other users' posts, even though it is completely voluntary. The Miner's goal is to correctly validate posts and remove malicious miners. There will be an organization to manage the rules and protocols of social media. The organization only deals with the management of social media, like hosting, adding new features, and other rules. This organization defines sets of rules and regulations even though it has no control over aspecific group or post.

VI. ALGORITHM

A. Consensus Algorithm

Every user will have some power. His or her power increases when they post good content (based on upvotes), and their power decreases when they post malicious content (based on downvotes and reports). We are using a modified version of Delegated Proof of Stake (DPOS). In each group, an election is conducted at specific time intervals. In each group, 20% of people compete in elections (based on the value of power they hold), and the rest of the people vote for witnesses for their groups. Our algorithm should take care that the same people do not become witnesses many times in a row.

B. Incentives Algorithm

The fund is generated through the cryptocurrency market and ads. There will be mainly two types of incentives: cryptocurrency rewards and power. When a user posts good content and gets upvotes, he can withdraw the incentive in the form of both cryptocurrency reward and power (e.g., 50% cryptocurrency reward and 50% power, or 100% power). Cryptocurrency rewards can be directly converted into fiat currencies for spending. Power adds more weight to the user's profile. But power is not directly convertible into fiat currencies. A user with higher power can participate in elections and get higher rewards for his or her posts in the future. When a user chooses to keep power instead of a cryptocurrency reward. He or she is indirectly investing in the cryptocurrency blockchain, which helps to increase the price of cryptocurrency. So, it is profitable for both users and the cryptocurrency market.

C. Revenue Model

Revenue is generated from ads and cryptocurrency trading. As users post good content, they receive cryptocurrency or power. When they receive power as a reward, they are indirectly contributing to the rise in the value of cryptocurrency. So, it is profitable for both users and cryptocurrency traders.

D. Solving Copyright Issues

When a user notices that his content is posted by another user, he will report that post for copyright infringement with a detailed report and proof of his original work. This report will be reviewed by multiple witnesses, depending on the power of the copyright claimant. The first step is the verification of digital signatures by using the user's public key. If the content is found to be copyrighted, the user's power will be suitably decreased, and that post will be taken down. If the copyright claim is false, the claimant's power will be suitably decreased. We assume that more than 50% of witnesses are honest and evaluate correctly. Even if a group of malicious witnesses tie up and perform malicious activities, they will be taken down in the next election by users, and the power of malicious witnesses will be suitably decreased. A newly created user account will have zero or negligible power. So, creating multiple accounts will not give any profit to the user since all of them will have zero power. If spamming is detected by witnesses, it will decrease the power of spammers. So, a user who has made efforts to gain power will not tend to do it. A user with zero power spamming will have no effect, as those posts are considered to have very little weight.

VII. RESULTS AND DISCUSSIONS

Copyright infringement, when raised by users, can be solved by miners; if miners are biased and malicious, they can be eliminated in the DPOS's next election and different miners elected. Assuming that more than 50% of the network is non-malicious, the risk of malicious users taking over the network is prevented. Users are rewarded with cryptocurrency for their honest actions, which in turn benefits social media as well. Instead of storing all the data in the blockchain like some of the current decentralized social media, which affects the performance as blockchain databases are not as fast as modern databases, only the data that needs to be preserved immutably is stored in the blockchain, and the rest of the content is stored in the modern database.

VIII. CONCLUSION

Copyright issues are increasing with the development of technology and the use of social media. We provide a solution to this issue with a blockchain-based approach. This idea can be successful if a large number of people start using decentralized social media because we need a large number of legitimate users to nullify the effect of malicious users.

IX. REFERENCES

[1]. Peng,Li&Feng,Wei&Yang,Laurence, Social-Chain: Decentralized Trust Evaluation Based onBlockchaininPervasiveSocialNetworking.ACMTransactionsonInternetTechnology.21.1-28.10.1145/3419102.(2021)

[2]. Murimi,Renita, ABlockchainEnhancedFrameworkforSocialNetworking, Ledger 4.10.5195/ledger. (2019).

[3]. Chakravorty, Antorweep & Rong, Chunming. Ushare:usercontrolledsocialmediabasedonblockchain.10.1145/3022227.3022325 (2017).

[4]. Mohamed Baza, Mahmoud Nabil, MohamedElsalihAbdelsalamMahmoud,NiclasBewermeier,KemalFidan,WaleedAlasmary,andMohamedAbdallah, Detecting Sybil attacks using proof ofwork and location in vanets. IEEE Trans. Depends. Sec.Computer Xiv:1904.05845v1. (2019)

[5]. ChristianDeckerandRogerWattenhofer, Informationpropagationinthebitcoinnetwork.InProceedingsoftheIEEEP2PConference.IEEE,1–10(2013)

[6]. YossiGilad,RotemHemo,SilvioMicali,GeorgiosVlachos,andNickolaiZeldovich.2017.Algorand:Scaling byzantine agreements for cryptocurrencies. InProceedingsofthe

[7]. 26thSymposiumonOperatingSystemsPrinciples.ACM,51–68.

[8]. NasreenAnjum,DmytroKaramshuk,MohammadShikh-Bahaei,and Nishanth Sastry, Survey onpeer-assistedcontentdeliverynetworks.ComputerNetworks116(2017).

[9]. Jan Trienes, Andrés Torres Cano, and Djoerd Hiemstra, RecommendingUsers:WhomtoFollowonFederatedSocialNetworks.(2018).arXiv:1811.09292

[10]. M.Ali,J.Nelson,R.Shea,andM.J.Freedman.Blockstack:Aglobalnamingandstoragesystemsecuredbyblockchains.In2016USENIXAnnualTechnicalConference(USENIXATC16), 2016.

[11]. SavvasZannettou,BarryBradlyn,EmilianoDeCristofaro,HaewoonKwak, MichaelSirivianos,Gianluca Stringini, and Jeremy Blackburn, whatis Gab: A bastion of free speech or an alt-right echochamber.InWWWCompanion. (2018)

[12]. F.Glaser,"Pervasivedecentralizationofdigitalinfrastructures:aframeworkforblockchainenabledsystem and use case analysis," in Proceedings of the50thHawaiiInternationalConferenceonSystemSciences,2017.

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Development of Medical Device Application using QT/QML

Dr.I.Manimozhi[1], Ms.Neha[2]
[1]Associate Professor, [2]Assistant Professor
Department of Computer Science & Engineering, East Point College of Engineering & Technology, Banglore, Karnataka, India

ABSTRACT
The main objective here is to provide a solution for Post Development Automation for Medical Device applications. Image-based testing can help in solving problems observed while automating QT/QML applications wherein objects may not be accessible through locators. This solution can help the automation to step ahead and provide accurate results on medical devices on executing the scripts. Algorithms need to be generated for Image-based testing In this paper we will discuss the search of Sub Images in the Main Image with Scalability, Color Resolution, and partly excluded images. Images are not usually named with convenient text or element ID and it's displayed as the image, so the photos are needed to be searched by their pixel. It is essential to compare reference images with the main image and to locate the pattern which we are looking for it. The pattern can be a picture of anything Number, Character, logo, building, or just a duplicate of the same image, and the search operation is to retrieve matching images with different option (scale and color).
Keywords: QT/QM, Post Development Automation Medical Device Application

I. INTRODUCTION

The main objective is to identify the sub image from the main image which is captured as screenshot from the device .This sub image can be further used to validate tests. So firstly we need to design an algorithm for the same. Further based on Image, Image Resolution, Image Size, partly excluded images we can design algorithms respectively. Based on exclude image areas Adding excluded areas to the reference image and only part of the image should be compared with captured image in the screen. In real time scenarios we may come across that a particular image say for example, a submit button might remain same but only the text inside it may vary according to languages chosen by the user. In such situations, we can exclude the text part and take the remaining portion which can be compared with the referenced image. So a solution needs to be designed such that the captured image excluding a specific area can be used to carry out the comparison with the reference image.

Based on Image Resolution: Resolution of images may vary from device to device based on the quality, color tolerance, screen resolution etc. The quality and color tolerance can be handled efficiently by including tolerance values, so that the images captured as screenshots can be compared with the reference images and

provide an accurate result. So here our aim is to provide a one end solution to it by designing an algorithm for the same.

Based on Image Size: Images can vary in size when captured as screenshots in different devices. The images can be either bigger or smaller in size

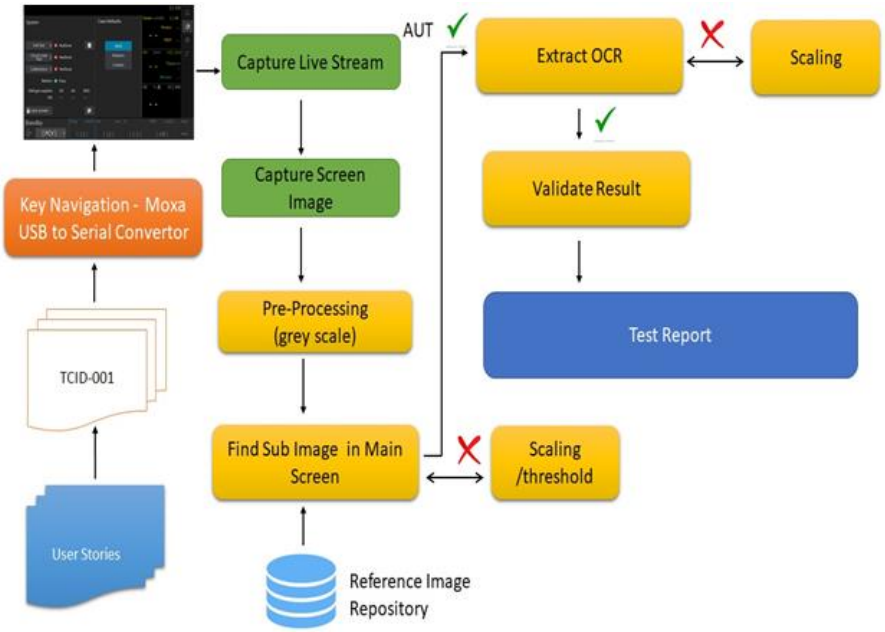
- Larger Image: An algorithm needs to be designed so that larger images that are captured as screenshots satisfies the condition when compared with the reference images during validation.
- Smaller Image: An algorithm needs to be designed so that smaller images that are captured as screenshots satisfies the condition when compared with the reference images during validation.

II. LITERATURE REVIEW

Background: Testing of Medical Device Applications (QT/QML) using Squish Automation tools helps to build test scripts which provides an accurate and efficient way of logging the test results. During development stage most Automation testing tools copying plugin/agents/packages in the targeted systems to run the test scripts But post development copying agents file into targeted systems is restricted due to security reason.

Problem Statement: To run automation and identify the regression issues during post development environment is challenging part in the medical devices. Third party software packages are restricted to copying in the medical devices during the post development and unable to run the automation scripts without third party plugin or agents in the QT/QML based targeted setup environment.

III. SYSTEM ARCHITECTURE



Approach: Image and text recognition is the backbone of automating the desktop applications. Image recognition is finding one image within another image. Reference image defined during the wireframe design time and actual image which is capture as a screenshot of the Real time application when the automation flow is running and validated against with reference image. Image based validation is the visual verification of rendering image on the screen for GUI application and its suitable for medical application.

Visual testing in medical software programs is a exceptional assurance activity for visible aspects of the utility's consumer interface and it's additionally referred to as Visual Validation Testing. Verifying the best information and content material are getting displayed on the utility's the front stop. Additionally, it also validates the format and appearance of each visual element present on the consumer interface and the complete UI itself.
Image search in the screenshot and extract the text, compare with reference text then follow the validation. Image with icon or object based search will not accurate based on resolution, color, image quality etc.. in ordered to avoid the complexity. text area image search algorithm required to avoid problems.

IV. IMPLEMENTATIONS

Algorithm to find Sub Image on Main Image

Consider a screenshot captured from a device as shown in "Fig 1". We need to identify a sub image "Fig 2" from the captured screenshot.

Algorithm:

Step 1: Begin:

Step 2: Input the source image (Original image)

Image src = ImageIO.read (new
File ("srcfile.jpg")

2.1: Define value of x- coordinate
Define value of y – coordinate
Define height
Define width

Step 3: Input the destination image file wherein the cropped image needs to be copied.

Step 4: Paint the desired part of the source image (src) onto the Destination image file (dst) using Graphics object

Step 5: Write the buffered image out from the Destination image file to a new file using ImageIO.

ImageIO . write (dst ,"png", new File("final_dst_file")

Step 6: Stop

Obtain text area image from the wireframe with scaled 1:1 - Load image, convert to grayscale, Gaussian blur, and adaptive threshold

Simple Thresholding became defined with extraordinary kinds of thresholding techniques. Another Thresholding technique is Adaptive Thresholding. In Simple Thresholding, a international fee of threshold was used which remained regular during. So, a steady threshold cost won't help within the case of variable lights conditions in different areas. Adaptive thresholding is the approach in which the brink value is calculated for smaller areas

cv2.ADAPTIVE_THRESH_MEAN_C: Threshold Value = (Mean of the neighbourhood area values – constant value). In other words, it is the mean of the blockSize×blockSize neighborhood of a point minus constant.

cv2.ADAPTIVE_THRESH_GAUSSIAN_C: Threshold Value = (Gaussian-weighted sum of the neighbourhood values – constant value). In other words, it is a weighted sum of the blockSize×blockSize neighborhood of a point minus constant.

Combine adjacent text. We create a rectangular structuring kernel then dilate to form a single contour

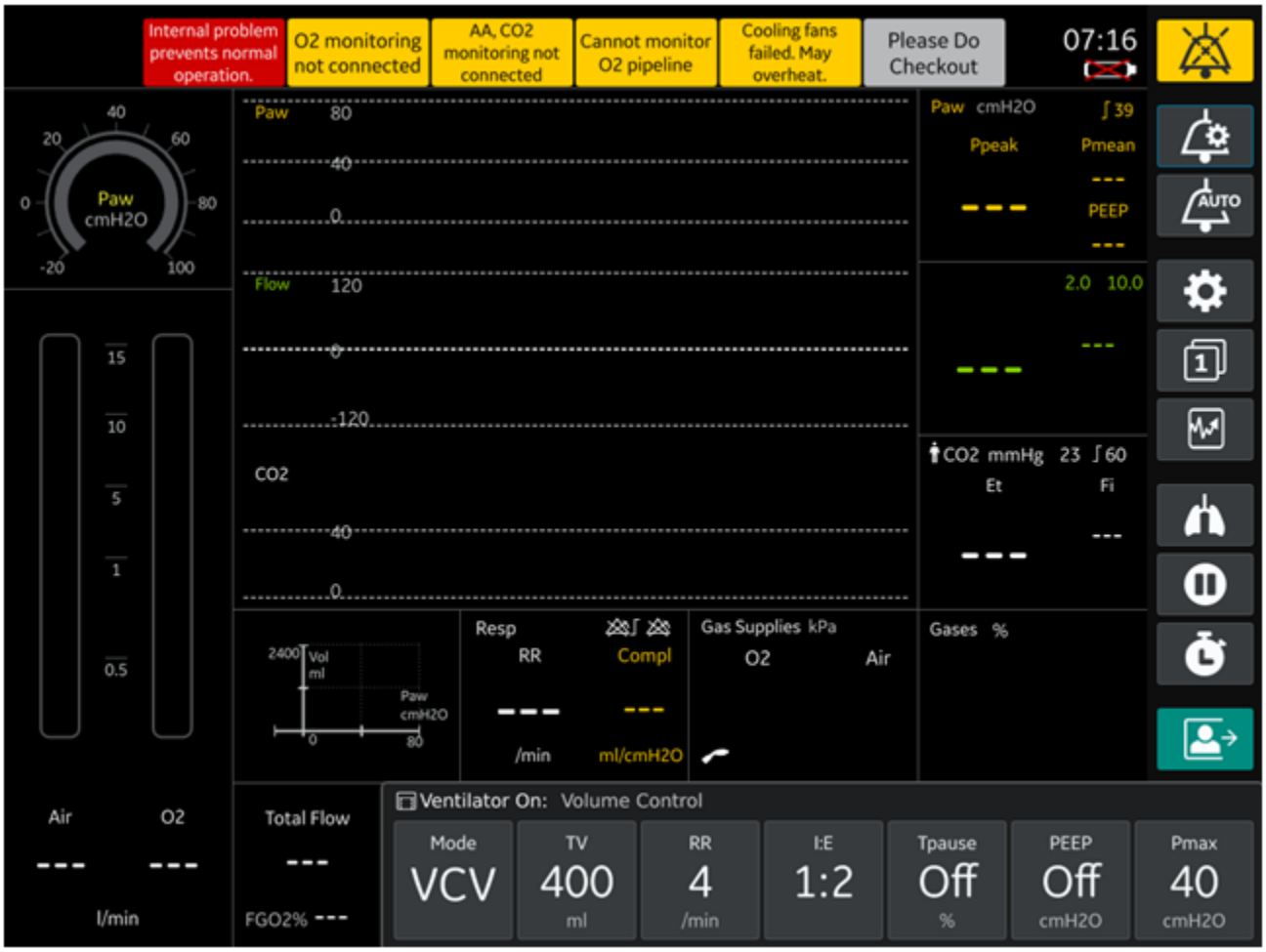


Fig 1: Image captured as screenshot from Wireframe Design

Algorithm to Upscale/Downscale of Reference image : Images needs to be upscaled/Downscaled according to Main image the pixels can be calculated and the exact position of the image can be located.

Algorithm:

Step1: Begin:

Step2: Input Source image (src)

2.1: Input the Captured Image (Screenshot from the device ass source Image)

Step3: Input Reference image to fins the sub image in the source image

Step4: Sub-image not detected then rescaling factor should be carried either upscaled /downscaled.

Step5: Search sub-image with incremental/decremental with scale factor

Step6: Re-Search Sub-Image – loops continues till target found

Step 8: Stop

Width and height of the image measured with help of sub image detection in the main image followed with Key navigation will be occurred. Again validation reference image search in the main image and extract text and validated with expected result .

Upscaled and downscaled mechanism will be taken both image search and OCR extraction

Automation Flow Chart:

Algorithm for dynamic text based on exclude image areas: Consider an image which remains fixed always and only a sub image of this image might vary frequently. In such situations the sub image can be cropped and the excluding area can be taken as the reference image to compare with the image obtained during real time.

We design an algorithm such that the sub image is detected from the original image and further that part is cropped from the original image. The image obtained after this is the reference image which is used for comparison.

Procedure1: Image cropping

Step 1: Begin:

Step 2: Input the source image (Original image)

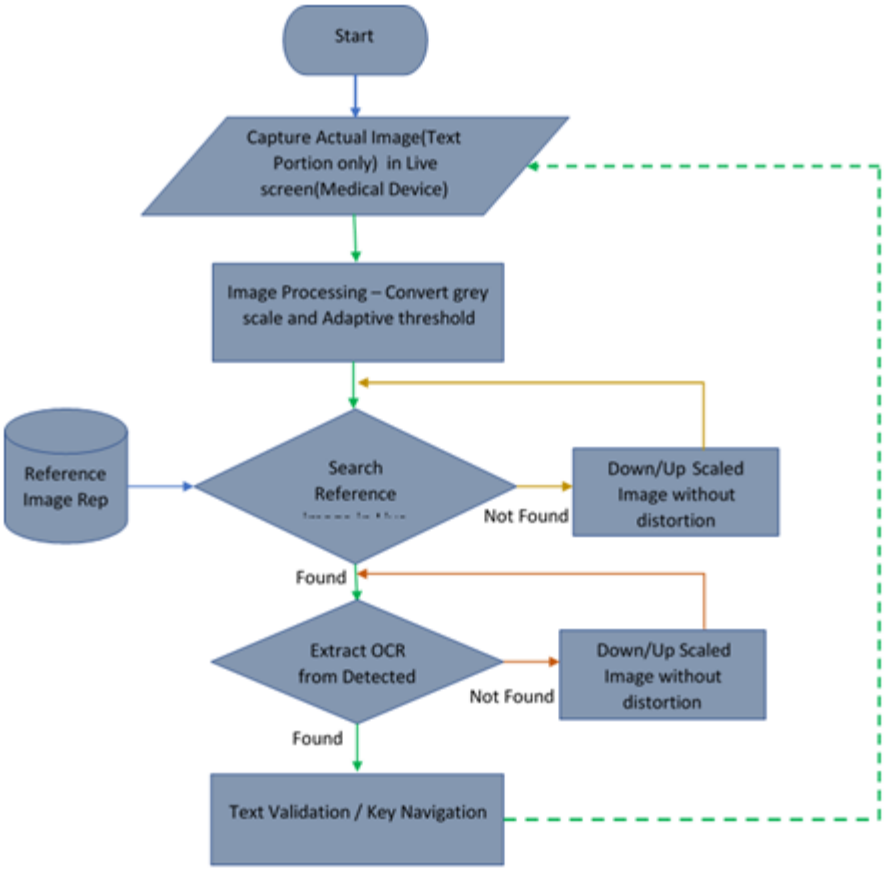
Image src = ImageIO.read (new File ("srcfile.jpg")

2.1: Define value of x- coordinate

Define value of y – coordinate

Define height

Define width



Algorithm:

Step 3: Input the destination image file wherein the cropped image needs to be copied.

BufferedImage dst= new BufferedImage (w, h, BufferedImage.TYPE_INT_ARGB)

Note: The width and height should be same as the source image

Step 4: Paint the desired part of the source image (src) onto the Destination image file (dst) using Graphics object dst.getGraphics().drawImage(src, 0,0,w,h,x,y,x+w,y+h,null)
Step 5: Write the buffered image out from the Destination image file to a new file using ImageIO.
ImageIO . write (dst ,"png", new File("final_dst_file")
Step 6: Stop
Procedure 2: Locate the image from the screenshot captured.
Algorithm findSubimage (BufferedImage Im1, BufferedImage Im2)
Step 1: Begin
Step 2: Input Im1 that is the screenshot captured from the Medical device
Input Im2 that is the cropped image "final_dst_file" obtained in Procedure1
2.1: Get the width and height of Im1 and Im2
w1 ← Im1. getWidth ()
h1 ← Im1. getHeight ()
w2 ← Im2. getWidth ()
h2 ← Im2. getHeight ()
assert (w2 < = w1 && h2 < = h1)
Step 3: Keep track of best position found
bestX ← 0
bestY ← 0
lowestDiff←Double.POSITIVE_INFINITY
Brute force search method through the whole image
for (int x =0 ; x < w1-w2 ; x++) do
for (int y =0 ; y < h1-h2 ; y++) do
double comp ← compareImages(Im1.getSubimage(x,y,w2,h2),Im2)
if (comp < lowestDiff) then
bestX ← x
bestY ← y
lowestDiff ← comp
print (lowestDiff)
return new int[] {bestX, bestY}

Step 4 : Determine how different two identically sized regions are Function compareImages (BufferedImage Im1, BufferedImage Im2)
assert(Im1.getHeight() == Im2.getHeight() && Im1.getWidth() == Im2.getWidth())
variation ← 0.0
for (int x = 0 ; x < Im1.getWidth() ; x++) do
for (int y = 0 ; y < Im1.getHeight() ; y++) do
variation += compareARGB(Im1.getRGB(x,y) ,Im2.getRGB(x,y)) / Math.sqrt(3)
return variation / (Im1.getWidth() * Im1 . getHeight())
end function
Step 5 : Calculate the difference between two ARGB Colours

```
function   compareARGB  (int rgb1 , int rgb2)
        r1 ← ((  rgb1 >> 16) & 0XFF)/255.0
        r2 ← ((  rgb2 >> 16) & 0XFF)/255.0
        g1 ← ((  rgb1 >> 8) & 0XFF)/255.0
        g2 ← ((  rgb2 >> 8) & 0XFF)/255.0
        b1 ← ((  rgb1  &  0XFF) / 255.0
        b2 ← ((  rgb2  &  0XFF) / 255.0
        a1 ← ((  rgb1 >> 24) & 0XFF)/255.0
        a2 ← ((  rgb1 >> 24) & 0XFF)/255.0
return a1 * a2 * Math.sqrt((r1-r2) * (r1 – r2 ) + (g1 –g2 ) * (g1 –g2 ) + (b1 –b2 ) * (b1 –b2 ))

end function

Step 6:   Stop
end algorithm
```

V. RESULTS

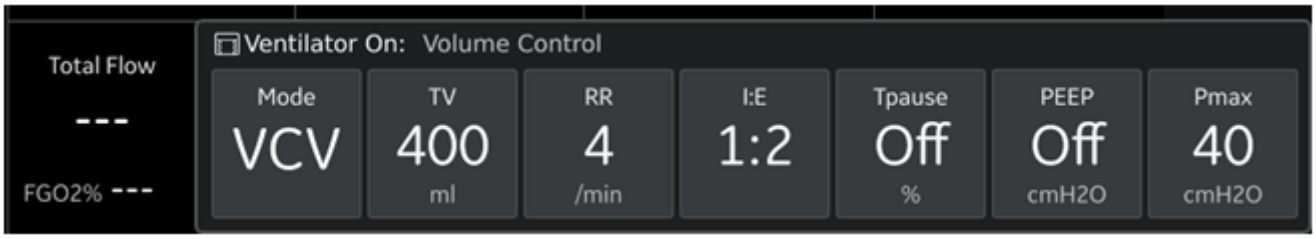


Fig 4: Screenshot obtained from the medical device application

Fig 4.1: Locate the Submit button from the screenshot in Fig 4 which is the value to be located with the text "400" being cropped off from the button (Src Image)

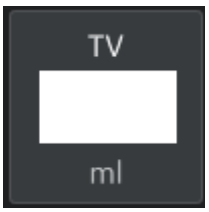


Fig 4.2: Submit button after cropping the text inside the button

Consider figure Fig 4 wherein the screenshot is captured from the medical device.

Consider the figure Fig 4.1 from which the Submit button whose text needs to be cropped off is located and then Consider figure Fig 4.2 which is the Submit button obtained after cropping of the text which is stored in the final_dst_file according to Algorithm in Procedure 1.

Procedure 2: The figure Fig 4.2 is compared with Fig 4.1 according to Algorithm of Procedure2 and thereby the subimage Fig 4.2 is located within the image in Fig 4.1

Algorithm based on Image Resolution: Color resolution of the medical application may vary from one device to another and therefore an algorithm should be designed to calculate the percentage of similarity between two images pixel wise. This percentage of similarity can be then compared with a tolerance value to judge the similarity

Algorithm:

compareImage (File fileA, File fileB)

Step1: Begin

Step2: Input the reference image file and the image file obtained from the screenshot which needs to be compared with the reference image.

Step3: Initialize the percentage value to

Float percentage = 0

3.1: Initialize the threshold value

3.2: Initialize the tolerance value

3.3: Calculate buffer data from both image files and store the value in a integer variable for further calculation

```
BufferedImage BufIm_A = ImageIO.read(fileA)
DataBuffer  DbufIm_A =  BufIm_A.getData().getDataBuffer()
Integer sizeA = DbufIm_A.getSize()
BufferedImage BufIm_B = ImageIO.read(fileB)
DataBuffer  DbufIm_B =  BufIm_B.getData().getDataBuffer()
Integer sizeB = DbufIm_B.getSize()
```

Step 4: Initialize a count variable count to 0

4.1: Compare data objects

```
If (sizeA == sizeB)
    for (int i = 0;          i < sizeA ; i++)
        if (DbufIm_A.getElem(i) ==
          (DbufIm_B.getElem(i) )
          count  += 1
        endif
    end for
   percentage = ( count  * 100 )/sizeA
 Else
   "Images are not of same size"
Return percentage
```

Step 5: Compare the percentage of similarity with the tolerance value

```
If (percentage <= threshold)
```

"Images are similar"
Else if (percentage + tolerance <=threshold)
"Images are similar"
Else if (percentage - tolerance <=threshold)
"Images are similar"
Else
"Images are not similar"

Step 6: End

Fig 5. Screenshot from Device

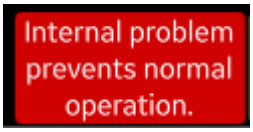


Fig 5.1. Reference Image

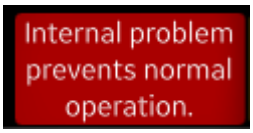


Fig 5.2. Image to be compared from Device

According to the Algorithm Figure Fig 5.1 and Figure Fig 5.2 are compared and based on the percentage of similarity the Images are found to be similar or not.
the image from the screenshot.

VI. CONCLUSION

The primary aim of this paper presentation was to find a solution to execute regression issues and layout verification with Automation scripts in medical device application using Image based Validation. We have listed down the scenarios that arise while automating and have tried to figure out solutions to the same. The idea presented here to execute the test scripts in post development environment without copying any third-party files/packages/agents in the targeted devices.

VII. REFERENCES

[1]. http://tech-algorithm.com/articles/nearest- neighbor-image-scaling
[2]. Richard Szeliski (Microsoft Research). Computer Vision: Algorithms and Applications szeliski.org/Book

[3]. Digital Image Processing Algorithms and Applications by I.Pitas
[4]. Principles of Digital Image Processing Burger, W. (et al.) (2019)
[5]. Saugat Bhattacharyy Kunal Pal "IoT-Based Applications in Healthcare Devices published in Journal of HealthCare Engineering, Hindawi, 2021
[6]. Jennifer C Goldsack " Verificaton,Analytical.Validation and Clinical Validation" published in Digitl Medicine.
[7]. A Disciplined Approach to User Interface Development Using Qt (QML) for Medical Devices, published in Medcuity software2020
[8]. S M Riyazul "The Internet of Things for Health Care: A Comprehensive Survey" published in IEEE access 2018

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

A Detection Tool for Finding Software Vulnerabilities in JAVA Code

Dr. Chandramouli H[1], Dr. I Manimozhi[2], Kesavan M V[3]
[1]Professor, [2]Associate Professor, [3]Assistant Professor
Department of CSE, East Point College of Engineering and Technology, Bengaluru, Karnataka, India

ABSTRACT

A vulnerability is a weakness in software that enables an attacker to compromise the integrity, availability, and confidentiality of the program and data that it processes. Software applications or programs are implemented in different languages and most of them contain serious vulnerabilities which can be exploited to cause security breaches. Many security vulnerabilities may be present in programs [1] and a number of techniques have been developed to detect these [2].

However, it is essential that these vulnerabilities are not only detected but corrected. Vulnerability management is the cyclical practice of identifying, classifying, remediating, and mitigating vulnerabilities [3]. The paper discusses a tool developed by the authors which not only detects software vulnerabilities but provides solutions for correcting them.

The tool also calculates the Degree of Insecurity in a Java program first defined in [2]. It applies the proposed mitigating methods and recalculates the Degree of Insecurity. Experimental results as discussed in the paper indicates that the mitigation methods suggested in the tool are very effective.

Keywords: Vulnerability, CWE, SecCheck, Mitigation, Degree of Insecurity (ISM).

I. INTRODUCTION

It is essential that software is engineered so that it continues to function correctly under malicious attacks. Software security is the process of designing, building, and testing software for security: it needs to identify and mitigate weaknesses so that software can withstand attacks. Many recent security breaches have been found to be due to inadequate design and improper coding. The vulnerabilities have unwanted consequences such as hijacking of session information, deletion or alteration of sensitive data, and execution of arbitrary code supplied by hackers [1].

It is essential that software vulnerabilities are not only detected but techniques used for correction of such vulnerabilities in order to prevent attacks and minimize operational disruption due to these. While Vulnerability Management [3][4][5] is being discussed by researchers, there are very few tools which actually does this.

The tool SecCheck developed by the authors can detect a number of vulnerabilities and has been described in [2] while the tool discussed here not only detects vulnerabilities in Java programs but also guides the developers by offering possible solutions for mitigating these. The tool detects vulnerabilities in any Java program caused by Null Point Dereference, Reachable Assertion, Allocation Of Resources Without Limits Or Throttling, Improper Validation of Integrity Check Value, Serializable Class Containing Sensitive Data, cleartext transmission of sensitive information, Improper Validation of Certificate With Host Mismatch, Improper Encoding or Escaping of Output, Improper Neutralization of CRLF Sequences in HTTP Headers, Use of Non-Canonical URL Paths for Authorization Decisions. It also offers solutions for mitigating these.
In Section 2 the pros and cons of the presence of weaknesses are discussed along with ways of mitigating these. Section 3 discusses Degree of Insecurity in a code. Section 4 discusses about the working of the tool while Section 5 describes results of experiments conducted to analyse the effectiveness of the tool.

II. COMMON VULNERABILITIES IN JAVA CODE

Common Software Vulnerabilities that occur in Java programs as discussed in CWE [1] are:

1. Reachable Assertion
2. Throttling
3. Null Pointer Dereference
4. Use of a One-Way Hash without a Salt
5. Use of Insufficiently Random Values
6. Missing Support for Integrity Check
7. Improper Neutralization of CRLF Sequences in HTTP Headers
8. Use of Non-Canonical URL Paths for Authorization Decisions
9. Reliance on Cookies without Validation and Integrity Checking in a Security Decision
10. Authentication Bypass by Spoofing
11. Unrestricted Upload of File with Dangerous Type
12. Improper Encoding or Escaping of Output

These vulnerabilities are detected in any Java program by the tool developed by the authours and solutions for mitigating these are offered. These are briefly discussed below along with the consequences and mitigations .

2.1. REACHABLE ASSERTION

Reachable assertion occurs when assert is triggered by an attacker causing crash of application or cause a denial of service[6].
While assertion is good for catching logic errors and reducing the chances of reaching more serious vulnerability conditions

Pros and Cons

1. Chat client allows remote attackers to cause a denial of service (crash) via a long message string when connecting to a server, which causes an assertion failure
2. Product allows remote attackers to cause a denial of service (crash) via certain queries, which cause an assertion failure

2.2. THROTTLING

The software allocates a reusable resource or group of resources on behalf of an actor without imposing any restrictions on how many resources can be allocated.

When allocating resources without limits, an attacker could prevent other systems, applications, or processes from accessing the same type of resources [7].

Pros and Cons

1. Driver Large integer value for a length property in an object causes a large amount of memory allocation
2. CMS does not restrict the number of searches that can occur simultaneously, leading to resource exhaustion
3. Product allows attackers to cause a denial of service via a large number of directives, each of which opens a separate window

2.3. NULL POINTER DEREFERENCE

Null pointer dereferencing occurs when a variable bound to the null value is treated as if it were a valid object reference and used without checking its state. This condition results in a NullPointerException, which can in turn result in a denial of service [8].

Pros and Cons

1. NULL pointer dereferences usually result in the failure of the process unless exception handling (on some platforms) is available and implemented. Even when exception handling is being used, it can still be very difficult to return the software to a safe state of operation.
2. In very rare circumstances and environments, code execution is possible.

2.4. USE OF A ONE-WAY HASH WITHOUT A SALT

The software uses a one-way cryptographic hash against an input that should not be reversible, such as a password, but the software does not also use a salt as part of the
input [9].

In cryptography, a **salt** is random data that is used as an additional input to a one-way function that hashes a password or passphrase.

The primary function of salts is to defend against dictionary attacks versus a list of password hashes and against pre-computed rainbow table attacks.

Pros and Cons

1. If an attacker can gain access to the hashes, then the lack of a salt makes it easier to conduct brute force attacks.
2. While it is good to avoid storing a cleartext password, the program does not provide a salt to the hashing function, thus increasing the chances of an attacker being able to reverse the hash and discover the original password if the database is compromised.

2.5. USE OF INSUFFICIENTLY RANDOM VALUES

The software may use insufficiently random numbers or values in a security context that depends on unpredictable numbers.

Pseudo-random number generators can produce predictable numbers if the generator is known and the seed can be guessed [10]. A random number generator (RNG) is a computational or physical device designed to generate a sequence of numbers or symbols that lack any pattern, i.e. appear random

Pros and Cons

1. When software generates predictable values in a context requiring unpredictability, it may be possible for an attacker to guess the next value that will be generated.
2. And use this guess to impersonate another user or access sensitive informat ion

2.6. MISSING SUPPORT FOR INTEGRITY CHECK

Checksumming is a well known method for performing integrity checks [11]. If the computed checksum for the current data input matches the stored value of a previously computed checksum, there is a very high probability the data has not been accidentally altered or corrupted.

Pros and Cons

1. Data that is parsed and used may be corrupted
2. Without a checksum it is impossible to determine if any changes have been made to the data after it was sent
3. Attackers can gain access to the sensitive information and can alter the data

2.7. IMPROPER NEUTRALIZATION OF CRLF SEQUENCES IN HTTP HEADERS

CRLF Injection is a software application coding vulnerability that occurs when an attacker injects a CRLF character sequence where it is not expected. Exploits occur when an attacker is able to inject a CRLF sequence into an HTTP stream [12].

CRLF Injection vulnerabilities result from data input that is not neutralized, incorrectly neutralized, or otherwise unsanitized.

Pros and Cons

1. CRLF Injection exploits security vulnerabilities at the application layer
2. Attackers can modify application data compromising integrity
3. Enables the exploitation of the following vulnerabilities:
    - XSS or Cross Site Scripting vulnerabilities
    - Proxy and web server cache poisoning
4. CR and LF characters in an HTTP header may give attackers control of the remaining headers and body of the response entirely under their control

2.8. USE OF NON-CANONICAL URL PATHS FOR AUTHORIZATION DECISIONS

The software defines policy namespaces and makes authorization decisions based on the assumption that a URL is canonical [13].

This can allow a non-canonical URL to bypass the authorization.Even if an application defines policy namespaces and makes authorization decisions based on the URL, but it does not convert to a canonical URL before making the authorization decision, then it opens the application to attack.

Pros and Cons

1. If a non-canonical URL is used, the server chooses to return the contents of the file, instead of pre-processing the file
2. An attacker can bypass the authorization mechanism to gain access to the otherwise-protected UR

2.9. RELIANCE ON COOKIES WITHOUT VALIDATION AND INTEGRITY CHECKING IN A SECURITY DECISION

The application uses a protection mechanism that relies on the existence or values of a cookie, but it does not properly ensure that the cookie is valid for the associated user [14].

Attackers can bypass protection mechanisms such as authorization and authentication by modifying the cookie to contain an expected value.

Pros and Cons

1. The cookie can be manipulated to claim a high level of authorization, or to claim that successful authentication has occurred

2.10. AUTHENTICATION BYPASS BY SPOOFING

This attack-focused weakness is caused by improperly implemented authentication schemes that are subjected to spoofing attacks.

An authentication mechanism implemented in java relies on an IP address for source validation. If an attacker is able to spoof the IP, however he may be able to bypass such authentication mechanism [15].

Pros and Cons

1. This weakness can allow an attacker to access resources which are not otherwise accessible without proper authentication

2.11. UNRESTRICTED UPLOAD OF FILE WITH DANGEROUS TYPE

The software allows the attacker to upload or transfer files of dangerous types that can be automatically processed within the product's environment [16].

Consequences

1. Arbitrary code execution is possible if an uploaded file is interpreted and executed as code by the recipient
2. The lack of restrictions on the size or number of uploaded files, which is a consumption issue

2.12. IMPROPER ENCODING OR ESCAPING OF OUTPUT

The software prepares a structured message for communication with another component, but encoding or escaping of the data is either missing or done incorrectly. As a result, the intended structured of the message is not preserved

If an application uses attacker-supplied inputs to construct a structured message without properly encoding or escaping, then the attacker could insert special characters that will cause the data to be interpreted as control information or metadata [17]. The component that receives the output will perform the wrong operations, or otherwise interpret the data incorrectly

Pros and Cons

1. The communications between components can be modified in unexpected ways
2. Unexpected commands can be executed, bypassing other security mechanisms. Incoming data can be misinterpreted

III. DEGREE OF INSECURITY IN JAVA CODE

Each of the weaknesses discussed in this paper has been assigned a severity level defined in CWE as shown in Table 1. We use a metric for calculating the Degree of Insecurity (referred to as ISM) [2].

$ISM = \sum_{i=1}^{m} Wi * N i$

where,

ISM stands for the Degree of Insecurity,

i is the Type of Vulnerability 1,2,....m

W_i is the Severity of Vulnerability in the software

Ni is the frequency of occurrence of vulnerability i.

Table 1: Severity of Vulnerabilities

Type of Vulnerability = i	Severity = Wi
Reachable Assertion	5
Throttling	4
Null Pointer Dereference	19
Use of a One-Way Hash without a Salt	12
Use of InsufficientlyRandom Values	15
Missing Support for Integrity Check	1
Improper Neutralization of CRLF Sequences in HTTP Headers	8
Use of Non-Canonical URL Paths for Authorization Decisions	1
Reliance on Cookies without Validation and Integrity Checking in a Security Decision	5
Authentication Bypass by Spoofing.	1
Unrestricted Upload of File with Dangerous Type	10
Improper Encoding or Escaping of Output.	6

IV. WORKING OF THE TOOL

The tool takes as input any Java program and scans to identify the vulnerabilities. If any vulnerability is detected then it displays warning message and suggests steps for its mitigation.

The steps followed are :

1. Select the input Java program
2. Select from the drop down list all types of vulnerabilities intended to be detected

As shown in Figure 1, for a Java program given as an input to the Tool, it displays type of vulnerability found and the place of its occurrence. It also gives the Degree of Insecurity in the input program

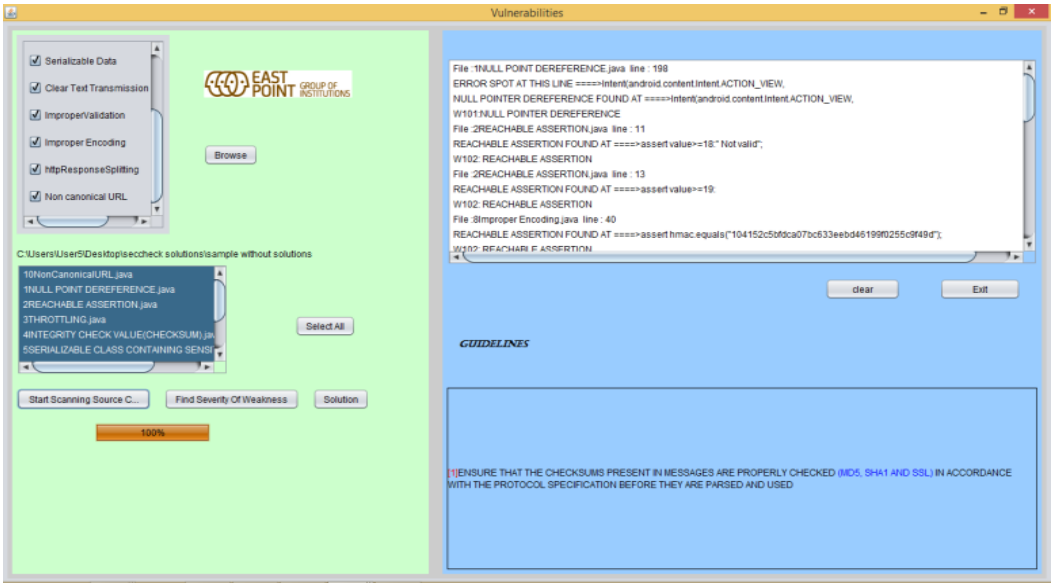


Figure 1: Front end of the Tool

The functional modules in the Tool are shown in Figure 2.*Scanner:* This module scans each line of source code one by one.

Pattern Matching Module: After scanning, the tool compares each line to find out if it contains a set of keywords which makes the program vulnerable to security threats. This is done by matching each line with the list of strings stored in a database.

Display Module: If there is a string match then a warning message is flagged to the user.

After the entire program is scanned, the *Degree of Insecurity* is calculated and displayed.

Mitigation Module: After the program is scanned, and Degree of Insecurity is found out, the next step is to mitigate the vulnerabilities detected. The tool provides solution for each of the vulnerabilities. The user has to incorpore the suggestions, so that the software becomes more secure.

Degreee of Insecurity is calculated again after this. If the Degree of Insecurity is acceptable, then the iteration stops. Otherwise, the steps are repeated and 'delinquent' statements are replaced.

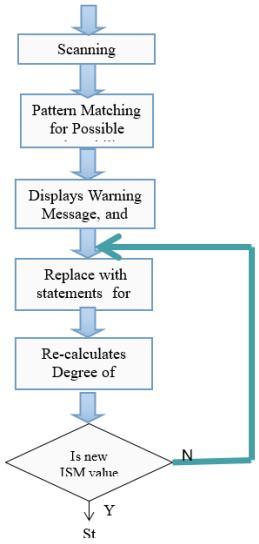


Figure 2: Working Scenario of the Tool

V. EXPERIMENTAL OUTCOMES

For detection of vulnerabilities by the tool we used programs written by the professionals taken from different vulnerability tracking sites and some were taken from Common Weakness Enumeration (CWE) site.
The results of measurements are given in Table 2. The vulnerabilities detected and the Degree of Insecurity Before Mitigation (BM) and After Mitigation (AM) are shown in these programs calculated as per the expression in Section 4 of this paper.

Table 2: Degree of Insecurity calculated in JAVA Code

Example name	Source	ISM (BM)	ISM (AM)
assert	http://www.javapractices.com/topic/TopicAction.do?Id=102	80	0
serverSocket	http://stackoverflow.com/questions/15541804/creating-the-serversocket-in-a-separate-thread	42	0
contentIntent	http://www.programcreek.com/java-api-examples/index.php?api=android.content.Intent.ShortcutIconResource (example 2)	95	0
hashmd5	http://howtodoinjava.com/2013/07/22/how-to-generate-secure-password-hash-md5-sha-pbkdf2-bcrypt-examples	60	0
httpurlconnection	http://www.mkyong.com/java/java-httpurlconnection-follow-redirect-example	60	0
UseDatagram	http://www.java2s.com/Code/Java/Network-Protocol/UseDatagramSockettosendoutandreceiveDatagramPacket.htm	22	0
Neutralization	http://stringpool.com/servlet-sendredirect-example	32	0
Non-Canonical URL Paths	https://code.google.com/p/crawler4j/source/browse/src/main/java/edu/uci/ics/crawler4j/url/URLCanonicalizer.java?r=b5b88a4d5c649a03e522b4e0557e7bbca1cc737b	48	0
Cookies	http://javabynataraj.blogspot.in/2011/04/what-is-deserialization-in-java-write.html	52	0
Authentication Bypass by Spoofing.	http://www.codereye.com/2010/01/get-real-ip-from-request-in-java.html	40	0
Unrestricted Upload	http://www.javacodegeeks.com/2013/08/servlet-upload-file-and-download-file-example.html	330	0

of File with Dangerous Type			
Improper Encoding or Escaping of Output.	http://www.programcreek.com/java-api-examples/index.php?api=javax.crypto.spec.SecretKeySpec (example 27)	18	0
Average Value of ISM Calculated from these Programs		**73.25**	**0**

VI. CONCLUSION

Vulnerability is a weakness in software. The causes of such "weakness" can be faults in design and in code and allows an attacker to reduce a system's information assurance.The presence of vulnerabilities in the software makes it necessary to have tools that can help programmers to detect and correct them during development of the code.

There are many tools available only to detect the vulnerabilities present in application programs written in various programming languages but no single tool has the capability to detect and mitigate the vulnerabilities found.

The tool developed by the authors and described in this paper detects and mitigates twelve vulnerabilities in Java source code.

VII. REFERENCES

[1]. http://www.cwe.mitre.org
[2]. Priyadarshini. R, Nivedita Ghosh and Anirban Basu "SecCheck: A Tool for Detection of Vulnerabilities and for Measuring Insecurity in Java Programs": International Journal of Software Engineering, Vol.7, No.2, July 2014, pp.67-93
[3]. Park Foreman, Vulnerability Management, Auerbach Publications, 2009
[4]. http://download.microsoft.com/download/5/0/5/505646ED-5EDF-4E23-8E84 6119E4BF82E0/Mitigating_ Software_ Vulnerabilities.pdf
[5]. A Agrawal and R A Khan, A Framework to Detect and Analyze Software Vulnerabilities -Development Phase Perspective", International Journal of Recent Trends in Engineering, Vol 2, No. 2, November 2009, pp 82-84
[6]. Reachable Assertion: "http://cwe.mitre.org/data/definitions/617.html"
[7]. Allocation of resources without limits or throttling: "http://cwe.mitre.org/data/definitions/770.html"
[8]. Null Pointer Dereference : "http://cwe.mitre.org/data/definitions/476.html"
[9]. use of a one-way hash without a salt: http://cwe.mitre.org/data/definitions/759.html
[10]. use of insufficiently random values: "http://cwe.mitre.org/data/definitions/330.html"

[11]. Missing support for integrity check: “http://cwe.mitre.org/data/definitions/353.html”
[12]. Improper neutralisation of CRLF sequences in HTTP headers:“http://cwe.mitre.org/data/definitions/113.html”
[13]. Use of Non-Canonical URL Paths for Authorization Decisions: “http://cwe.mitre.org/data/definitions/647.html”
[14]. Reliance on Cookies without Validation and Integrity Checking in a Security Decision: “http://cwe.mitre.org/data/definitions/784.html”
[15]. Authentication Bypass by Spoofing. http://cwe.mitre.org/data/definitions/290.html
[16]. Unrestricted Upload of File with Dangerous Type. http://cwe.mitre.org/data/definitions/434.html
[17]. Improper Encoding or Escaping of Output. http://cwe.mitre.org/data/definitions/116.html

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Fog Computing and its Current Research Trends

Rashmi T V*[1], Divya U H[1], Manimegalai A[1], Jagadevi Bakka[1]
*[1]Assistant Professor
Department of Computer Science and Engineering, Bangalore, Karnataka, India

ABSTRACT

The Internet of Everything (IoE) solutions gradually bring every object online, and processing data in a centralized cloud does not scale to the requirements of such an environment. This is because there are applications such as health monitoring and emergency response that require low latency and delay caused by transferring data to the cloud and then back to the application can seriously impact the performance. To this end, Fog computing has emerged, where cloud computing is extended to the edge of the network to decrease latency and network congestion. Fog computing is a paradigm for managing a highly distributed and possibly virtualized environment that provides compute and network services between sensors and cloud data centers. The computing world has seen a paradigm shift from traditional personal computing to present-day client-server computing with the advancements in computer networking. Client-server computing has completely evolved into cloud computing, which provides flexibility, low-cost deployment, fault tolerance, and high availability, to build virtualized services. Currently, with the proliferation of the Internet of Things (IoT) devices, computing needs latency-sensitive support, which a cloud cannot provide. In the year of 2012, a group of researchers from Cisco presented a new computing paradigm, called fog computing, where IoT devices can be given effective and enhanced support by bringing back a part of the computation from the cloud to the edge or near-edge devices. Fog computing is a computing paradigm where some of the computations take place in the edge devices, and these fog devices interplay with the cloud server to provide a better quality of service (QoS) to the end-users.

KEYWORDS: Fog Computing; IoT; Edge Computing; Cloud Computing

I. INTRODUCTION

Fog computing is a decentralized computing infrastructure in which data, computing, storage, and applications are located somewhere between the data source and the cloud. Like edge computing, fog computing brings the advantages and power of the cloud closer to where data is created and acted upon. Many people use the terms fog computing and edge computing interchangeably because both involve bringing intelligence and processing closer to where the data is created. This is often done to improve efficiency, though it may also be used for security and compliance reasons.

The fog extends the cloud to be closer to the things that produce and act on IoT data. These devices, called fog nodes, can be deployed anywhere with a network connection: on a factory floor, on top of a power pole,

alongside a railway track, in a vehicle, or on an oil rig. Any device with computing, storage, and network connectivity can be a fog node. Examples include industrial controllers, switches, routers, embedded servers, and video surveillance cameras.

II. CHARACTERIZATION OF FOG COMPUTING

Fog Computing is a highly virtualized platform that provides computing, storage, and networking services between end devices and traditional Cloud Computing Data Centers, typically, but not exclusively located at the edge of the network. Figure 1 presents the idealized information and computing architecture supporting future IoT applications and illustrates the role of Fog Computing.

Compute, storage, and networking resources are the building blocks of both the Cloud and the Fog. "Edge of the Network", however, implies several characteristics that make the Fog a non-trivial extension of the Cloud. Let us list them with pointers to motivating examples.

Edge location, location awareness, and low latency. The origins of the Fog can be traced to early proposals to support endpoints with rich services at the edge of the network, including applications with low latency requirements (e.g. gaming, video streaming, augmented reality).

Analyzing IoT data close to where it is collected minimizes latency. It offloads the gigabytes of network traffic from the core network. And it keeps sensitive data inside the network.

IoT environments consist of loosely connected devices that are connected through heterogeneous networks. In general, the purpose of building such environments is to collect and process data from IoT devices to mine and detect patterns, perform predictive analysis or optimization, and finally make smarter decisions promptly. Data in such an environment can be classified into two categories [10]

- Little Data or Big Stream: transient data that is captured constantly from IoT smart devices.
- Big data: persistent data and knowledge stored and archived in centralized cloud storage fog computing also have the potential of providing services as follows.

1) Location awareness: The fog device of a particular location can better know its context information.
2) Wide-spread-geographical distribution: The fog nodes are distributed around large geography.
3) Mobility-based services: Mobile devices can move with uninterrupted fog-enabled services.
4) Supporting a very large number of nodes: A large number of end devices can be served in the fog architecture.
5) Omnipotent role of wireless access: Wireless network has provided the advantage of accessing the fog services.
6) Device heterogeneity: Different heterogeneous devices can reside and participate in the fog computation with minimal effort.

III. FOG COMPUTING ARCHITECTURE

The concept of fog computing [4] emerged from the concept that part of the computing can be brought back near the edge devices. The term fog computing has been proposed by Cisco [2] in 2012

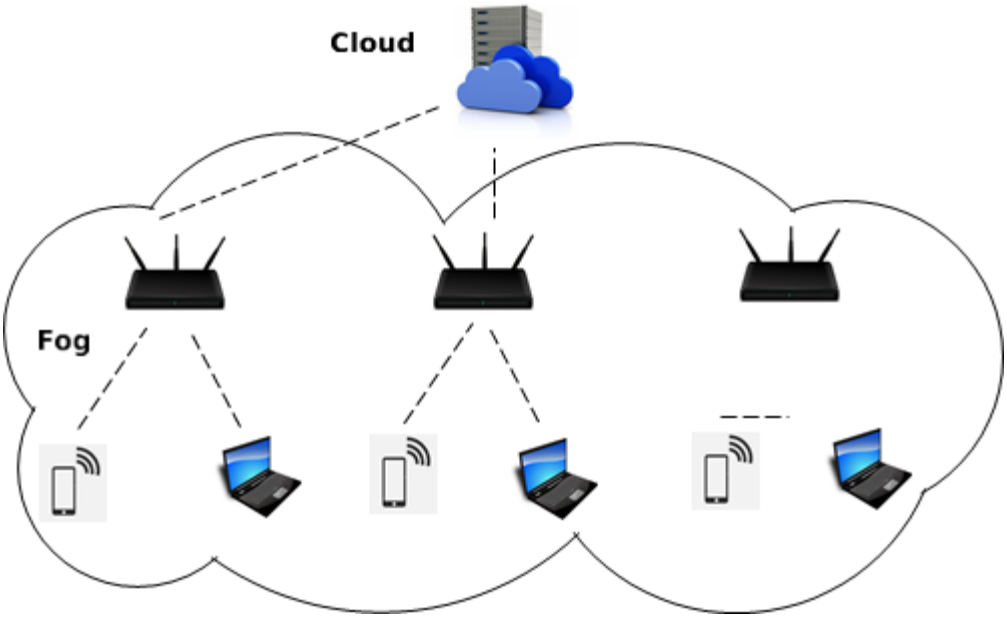


Fig. 1. Typical Components of a Fog Architecture

The fog computing architecture has been shown in fig. 1. It refers to extend the cloud computing paradigm to the edge of the network. The edge devices (i.e. routers, gateways, etc.) can be used as the computing nodes along with the existing cloud data centers. Fog computing has been envisioned to provide computation from the network edge, through the network core, and to the cloud data centers. The different services are hosted in the fog nodes, which are using its resources through the hypervisor, the management software for virtualizing the computing environment. Fog computing does the proper interplay of the services with the cloud. The applications which require real-time response and context-aware computing rely on the fog computing framework. Further, there are situations where there is a need for supporting a huge amount of data generated from IoT devices. Cloud computing alone is not sufficient in these situations as there is a requirement for real-time service provisioning. The typical applications of the fog computing paradigm can be in real-time healthcare monitoring systems, smart cities, smart grids, vehicular ad-hoc networks (VANET), etc. Being loosely coupled and highly distributed, QoS management and dynamic adaptability are the key challenges faced by the fog computing domain which need to be solved.

IV. FOG COMPUTING: RESEARCH CHALLENGE

Fog computing is a domain that emerged from the success of the cloud computing framework as a commercial and commodity solution for providing a computing resource to the end-users. However, with the development of low-cost computing hardware and devices like IoT sensors and smartphones, the research community realized that a part of the computation can be brought back to the devices near the edge, which can reduce the cost of data offloading at the cloud, as well as can provide privacy and security solution to the user data. However, computation at the edge also has challenges that the researchers are currently exploring for the end-to-end development of the fog computing framework. Consequently, several research outcomes have come out recently. In this section, we briefly discuss the general challenges for the development of a fog computing solution and accordingly classify the existing works into different groups for further discussion.

4.1. Challenges in Fog Computing Research

As mentioned earlier, fog computing is a distributed computing architecture that involves network-related challenges, computing-related research directions, security-related challenges, as well as management-related challenges. Being highly distributed, it makes the system more vulnerable to computation correctness. Here, we discuss these issues of a fog computing system.

Network and Device-Related Challenges: The various network and device-related challenges that the fog computing framework faces are as follows.

Distributed architecture: The distributed architecture makes fog computing more prone to having a redundant system. The same code is replicated in several locations in the edge devices of the network[3]. Therefore, the computing framework should have sufficient sophistication to reduce redundancy in the distributed environment.

Network resource distribution: The networking resources are scattered in the edge or near-to-edge devices in the fog architecture. This makes the system more complex in terms of the network connectivity aspects. A common network middleware is required to be developed, which can manage the common pool of resources over the edge or near-to-edge devices, and accordingly should be able to allocate resources to the application workloads.

Heterogeneity of devices: The fog environment has several heterogeneous end devices. This heterogeneity of the devices has made the system more diverse[5]. The computing platform should consider this device and network heterogeneity while developing fog applications.

Computing Challenges: The computation over a fog environment is challenging because of the following reasons.

Computation hierarchy development: The fog computing system always communicates with distant cloud servers. There is a trade-off between the response time and computation power in the fog computing system. The fog computing devices, that is the edge as well as the end devices, perform computations and should respond to the users within a time guarantee. At the same time, some computations are also offloaded to the cloud, and these computations at the cloud may take higher time compared to the time required to execute the computation at the edge devices but with a less computation cost. Therefore, it is always a challenge to identify what parts of the computations have to be offloaded to the cloud, and what fractions of the computations have to be performed at the fog devices. This also has the trade-off in accuracy-interoperability that needs to be addressed by the application developer.

Computation resource distribution: Computation of different applications needs proper resources. The edge devices may not always have all the resources deployed in them. Some of the resources have to be used from other fog nodes. This requirement has generated the need to distribute the computation resources among the edge devices. Therefore, there is a requirement for developing a converged framework to integrate the computation, memory as well as networking resources for building up the common pool. Applications can reserve resources from this common pool. The current research in this direction is exploring whether container technology [9] can be used to develop a common pool of resources over the edge devices for computation.

Distributed computation: The distributed computation in the fog has created the need to verify the computation's correctness. Fog applications need to be designed and developed in such a way that there are few inconsistencies in computation, and also such inconsistencies should be verifiable [3].

Mobility: With the advent of mobile and hand-held devices, the current computing framework demands computation over anywhere, anytime and anything connectivity, and therefore a pervasive computing framework needs to have emerged over the fog computing framework. The edge nodes may be mobile in the fog computing environment. This mobility is another barrier to computing in the fog domain. Therefore, the researchers need to develop integrated, pervasive, and ubiquitous solutions for handling the mobility of fog-enabled computing frameworks [8].

Security-Related Challenges: The fog computing system, being distributed with different heterogeneous devices, is vulnerable to various security attacks. The existing literature discusses man-in-the-middle attacks in the fog computing domain [8]. Data and network security are the main issues in fog. Further, as the fog computing framework also depends on the services from the cloud servers, the computation framework becomes vulnerable to trust and authentication issues. The privacy of the data is another concern in this highly distributed fog computing architecture[8].Another security vulnerability is that fog devices are not deployed in highly secure data centers, but in locations that may be easy to have physical access for attackers[8]. Hence, the system software itself may not be trusted. Therefore, there is a requirement to securely execute the edge functionalities over the fog.

Management Challenges: The fog computing framework, being a distributed system architecture, poses several challenges related to system management.

Service-oriented computing: In the fog computing framework, user service is divided into multiple micro-level services and these micro-services are distributed across the edge devices and the cloud. This particular distribution of services over fog devices is a mode of service-oriented computation over edge devices. However, executing micro-services over the fog nodes has its challenges. The proper management of the architecture to get the services is one of the prime challenges in the fog computing domain. There are several challenges in micro-service management. These are service placement, service combination, tracking of execution steps, etc. We need a proper orchestration system so that the services are provided to the end-users within a very less amount of time over the fog framework [8].

Resource management: The different networking, as well as computation resources, are distributed in the fog computing domain [8]. Fog computing has to be flexible and adaptive (like the cloud) to respond to issues like transient failures or resource shortages. The failure of fog nodes makes the whole system down as the resource would not be available from that fog node. Again, the resources are virtualized in the fog network. The virtualization of resources creates many challenges. These challenges are the latency, initiation, placement, migration of virtual network devices in a fog network, etc. In these cases, we need to properly manage the resources so that downtime can be avoided ensuring high availability. This is primarily because a fog computing system deals with latency-sensitive applications, such as smart homes, smart healthcare monitoring systems, etc. [3] [8]. Modern technologies like software-defined networking can be utilized for resource management [7] over the fog nodes which poses several research directions.

Orchestration between fog nodes and the cloud: Another challenge is the end-to-end orchestration of the fog-cloud resource ecosystem to provide QoS guarantees for various user-level services [8]. The fog computing system consists of the edge network as well as the cloud infrastructure. The integration of heterogeneous edge devices needs to be taken care of in the fog environment. Also, the cloud infrastructure should be properly handled to perform distributed computation and storage. Thus, there is a requirement for end-to-end orchestration of cloud servers and heterogeneous fog devices so that the resources can be allocated dynamically.

Based on these diverse challenges to develop an end-to-end fog framework, the researchers and industrial developers have targeted to solve various aspects of the fog computing framework. Accordingly, we classify the existing literature on fog computing, as discussed next.

The following are a few of the applications using fog computing.

Healthcare: An analytics system assisted by fog computing called FAST was proposed by Cao [6] can monitor fall conditions for stroke patients. They developed a set of fall detection algorithms, which uses data like measurements of acceleration and time series analysis methods along with data noise-reducing algorithms to allow an increase in efficiency in detecting fall conditions. It detects fall conditions in real-time and it was based on a distributed network of fog computing.

Augmented Reality: Applications based on augmented reality cannot allow even minor latency as even very small delays in response can potentially damage the user experience. Fog computing will be one of the major players in the augmented reality domain due to its distributed computing capabilities. The augmented Brain-Computer Interaction Game proposed by Zao [7] is based on Fog Computing and Linked Data, it will collect raw streams of data created by EEG sensors and then it will be classified to detect the brain state of the user while playing a game, which uses augmented reality [11]. Brain state classification is one of the most computationally heavy signal processing tasks, which needs to be carried out in real-time.

Caching and Preprocessing: Zhu[1] improves website performance by using edge servers where users used to connect with the internet by "fog boxes" where each user's HTTP request goes through a fog device. The fog device on user's page loading requests, to reduce its amount of time, performs various optimizations. It will have some general time-saving techniques like caching HTML components, reorganizing the composition of webpages as well as reducing the size of elements in the web. The edge devices could also perform different optimizations that will analyze the user's behavior and internet conditions. For example, when there is congestion in the network, the device at the edge then sends low-resolution graphics and photos to the user so that response time can be under an acceptable limit. Also, the performance of the client machines is monitored by the edge device which sends graphics of appropriate resolution taking care of the rendering time required by the browser.

4.2. Software-Defined Networks (SDN): SDN is an important concept based on computing and networking. SDN along with fog computing together will be capable to resolve some of the main issues related to networks of vehicles, intermittent nature of connectivity, high collision rate, and high rate of packet losses. Augmenting the vehicle-to-vehicle as well as vehicle-to-infrastructure communications along with main control supported by the SDN does this. It would split the control and communication layers where controlling will be done by the central server and the server would decide the communication path for nodes.

V. COMPARISON OF FOG COMPUTING AND CLOUD COMPUTING OVER IOT

Even though fog computing gives great advantages to the IoT infrastructure, however, cloud computing is one of the emerging solutions and it is already in use in many different areas. Also, a lot of research and development had already happened into cloud computing compared to fog computing. Fog and cloud are both good solutions but they are the complement of each other in the form of providing service. Table 1 shows a comparison of Fog and cloud computing.

Table 1- Comparison of Fog and cloud

Areas[8]	Cloud	Fog
Location and Model of Computing	Centralized in a small number of data centers	Distributed along large Geo-graphical areas and it is closer to the user. Fog nodes and systems can be controlled by a centralized node or in a distributed manner.
Size	High Each cloud data center is very large in size consisting of at least thousands of servers.	Each fog node can be equivalent to a single server machine. It's designed to meet user demands.
Deployment	Require sophisticated deployment planning	Depends on the environment. The majority of them don't require intense planning.
Operation	Operated in a fully controlled environment with technical expert teams by large companies.	Operated in an environment where it primarily depends on user demands. They are not operated directly by a person. It can be operated by any size of the company.
Applicatio-ns	It can support predominantly cyber-domain applications. The applications mainly suffer high latency.	Can support cyber-domain, and cyber-physical applications. It suffers very less latency and is hence useful for time-critical applications.
Network requireme-nts	Require clients to have network connectivity until the user wants to access its services. Bandwidth requirement grows with the increase in the total amount of data generated by all the clients.	Can operate autonomously to provide uninterrupted network services even with no or intermittent network connectivity. Bandwidth requirement depends on the total amount of data that needs to be sent to the cloud after filtering by fog.

VI. CONCLUSION

Fog computing has benefits in many domains and provides solutions for security problems in the current paradigm. Future research can expand the paradigm of fog computing in smart grids. This provides us a vision for the fog to be a platform of unification made enough for a brand new breed of rising services and modify the development of the latest devices and applications. However, this study suggests that with the emergence of fog, cloud computing can not become an obsolete concept. As it was seen, even if fog computing continues to develop in the future, it has to go hand-in-hand with cloud computing. Fog computing is just a bridge that connects the gap between the demands of IoT infrastructure and the computational capabilities of the cloud could provide. By creating harmony between these two complementary techniques of remote computing, various targets can be achieved, which are once just a dream for us.

VII. REFERENCES

[1]. Fog Computing: principles, architectures, and Applications by A.V. Dastjerdi, H. Gupta, R.N. Calheiros, S.K. Ghosh, R. Buyya Available online 17 June 2016, Version of Record 17 June 2016 - Chapter 4 (Fog Computing: principles, architectures, and applications - ScienceDirect)

[2]. Fog Computing: A Primer by Matthew N. O. Sadiku, Mahamadou Tembely, and Sarhan M. Musa, International Journals of Advanced Research in Computer Science and Software Engineering ISSN: 2277-128X (Volume-7, Issue-7)

[3]. A Distributed Fog Computing Architecture Supporting Multiple Migrating Mode by Qimin Jin; Rongheng Lin; Hua Zou; Fangchun Yang 2018 5th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2018 4th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)

[4]. Fog Computing State Of The Art: Concept And Classification Of Platforms To Support Distributed Computing Systems by Alexandra A. Kirsanova, Gleb I. Radchenko, Andrei N. Tchernykh

[5]. Fog Computing in the IoT Environment: Principles, Features, and Models by Somayya Madakam & Pratima Bhagat (Fog Computing in the IoT Environment: Principles, Features, and Models | SpringerLink)

[6]. Fog-based healthcare systems: A systematic review by Zahra Ahmadi, Mostafa Haghi Kashani, Mohammad Nikravan, and Ebrahim Mahdipour, Published online 2021 Sep 4(Fog-based healthcare systems: A systematic review - PMC (nih.gov))

[7]. Augmented Brain-Computer Interaction Based on Fog Computing and Linked Data by Sergio José Rodríguez Méndez, Australian National University John Kar-Kin Zao National Chiao Tung University Conference: 2014 International Conference on Intelligent Environments (IE) / Publisher: IEEE At: Shanghai, China

[8]. Security in fog computing: A systematic review on issues, challenges, and solutions Received 17 August 2020, Revised 18 June 2021, Accepted 5 August 2021, Available online 13 August 2021, Version of Record 13 August 2021

[9]. Adoption of community monitoring improves common pool resource management across contexts Published online 2021 Jul 12 (Sustaining the Commons: Advancing Understanding of Common Pool Resource Management Special Feature: Adoption of community monitoring improves common pool resource management across contexts - PMC (nih.gov))

[10]. Improving Web Sites Performance Using Edge Servers in Fog Computing Architecture byJiang Zhu, D. Chan, M. Prabhu, P. Natarajan, Hao Hu, F. Bonomi ,Published 25 March 2013, Computer Science 2013 IEEE Seventh International Symposium on Service-Oriented System Engineering

[11]. Can be get at https://ugcnetpaper1.com/sources-acquisition-and-classification-of-data/

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Framework for Efficient Resource Allocation in Cloud with Green Computing

Mrs. K. Revathi
Assistant Professor
Department of Computer Science and Engineering, East Point College of Engineering, Bengaluru, Karnataka, India

ABSTRACT
Green cloud depict the potential environmental benefits that information technology services delivered over the Internet can offer society. Green computing is the environmentally responsible and eco-friendly use of computers and their resources. This practices include the development of environmentally sustainable production practices, energy efficient computers and improved disposal and recycling procedures. In this paper we present a system where virtualization is used to allocate resources based on application demands. Here the main aim is to minimize the skewness. "Skewness" is uneven resource utilization of a server. It prevent overload in the system effectively by migrating the request from server to server. Because when the PM is overloaded it can lead to degraded performance of its VMs. Here a set a algorithms were used to predict the load and avoid overloading of Physical Machines. Idle Physical machines are turned off to save energy. Overload avoidance is maintained in this paper which leads to achieve good performance.
Keywords— Cloud Computing, Virtualization, Green Cloud

I. INTRODUCTION

The primary environmental problem associated with the cloud is energy use. According to market research conducted by Pike Research, the wide-spread adoption of cloud computing could lead to a potential 38% reduction in 2020. virtualization is software that separates physical infrastructures to create various dedicated resources. It is the fundamental technology that powers cloud computing. Virtualization software makes it possible to run multiple operating systems and multiple applications on the same server at the same time. Virtualization software makes it possible to run multiple operating systems and multiple applications on the same server at the same time. Virtual machines (VMs) provide flexibility and mobility through easy migration, which enables dynamic mapping of VMs to available resources.

Cloud computing has several appealing implications where under-provisioning of resources compromise service quality and over-provisioning wastes investment as well as electricity.Some times in data mining applications, a user may require a large number of servers for a short period of time.

Such requirement is possible only because of cloud computing.Main uses of virtualization concepts are:
Maximize resources — Virtualization can reduce the number of physical systems you need to acquire, and you can get more value out of the servers. Most traditionally built systems are underutilized. Virtualization allows maximum use of the hardware investment.
Multiple systems — With virtualization, you can also run multiple types of applications and even run different operating systems for those applications on the same physical hardware.
IT budget integration — When you use virtualization, management, administration and all the attendant requirements of managing your own infrastructure remain a direct cost of your IT operation.

II. RELATED WORK

The Green computing concept of closing the idle server is discussed in [6][7]. In *LiteGreen*[6], the desktop system saves energy by migration between user PM and VM server. Thus, the user's desktop environment is "always on", maintaining its network presence fully even when the user's physical desktop machine is switched off and thereby saving energy. This seamless operation allows LiteGreen to save energy during short idle periods as well the desktop VM is migrated back to the physical desktop machine. Thus, even when it has been migrated to the VM server, the user's desktop environment remains alive (i.e., it is "always on"), so ongoing network connections and other activity are *not* disturbed, regardless of the application involved.
According to dynamic server migration and consolidation algorithm is introduced. The algorithm is shown to provide substantial improvement over static server consolidation in reducing the amount of required capacity and the rate of service level agreement violations. Benefits accrue for workloads that are variable and can be forecast over intervals shorter than the time scale of demand variability. The management algorithm reduces the amount of physical capacity required to support a specified rate of SLA violations for a given workload by as much as 50% as compared to static consolidation approach. Another result is that the rate of SLA violations at fixed capacity may be reduced by up to 20%.
In Energy aware server provisioning and load dispatching for connection intensive internet services. Two techniques are used 1. dynamic provisioning that dynamically turns on a minimum number of servers required to satisfy application specific quality of service. 2. Load dispatching that distributes current load among the running machines.
According to sleep server concept SleepServer, a system that enables hosts to transition to such low-power sleep states while still maintaining their application's expected network presence using an on demand proxy server. Our approach is particularly informed by our focus on practical deployment and thus SleepServer is designed to be compatible with existing networking infrastructure, host hardware and operating systems. Using SleepServer does not require any hardware additions to the end hosts themselves, and can be supported purely by additional software running on the systems under management.

III. PROPOSED SYSTEM OVERVIEW

The User request for the VM and the VM scheduler checks for the authentication of request.

After that the skewness algorithm is performed and skewness is calculated. The result of skewness is forwarded to the VM scheduler. It checks the server status and also it verifies the status. If the server is not HOTSPOT, then the request is forwarded to server otherwise migration Process takes place.
Here the request is migrated from Hotspot server to warm spot server. So that Load balancing can be achieved. Then to achieve the Green computing the server which is in cold spot for long time a timer is triggered. The VM scheduler request for the server details and also its active connections. The server monitors the details and verify whether to shut down the server. The VM scheduler request for the prediction algorithm and its predicts the future resource needs. Then the processes running in that servers are migrated and the particular cold spot server is shutdown to achieve Green computing.

IV. OVERLOAD AVOIDANCE

In favour of performance and stability, we designed a pragmatic algorithm, Skewness. It is inspired from the fact that if a PM runs too many memory intensive VMs with light CPU load, much CPU resources will be wasted because it does not have a enough memory for extra VM. Here skewness is used to qualify the unevenness in utilization of multiple resources on a server.
The resource skewness of server p is given as

$$\text{skewness}(p) = \sqrt{\sum_{i=1}^{n} \left(\frac{v_i}{\bar{v}} - 1\right)^2}$$

Here n - number of resources
Vi - utilization of i th resource
By minimizing the skewness, we can combine different types of workloads nicely and improve the overall utilization of server resources.

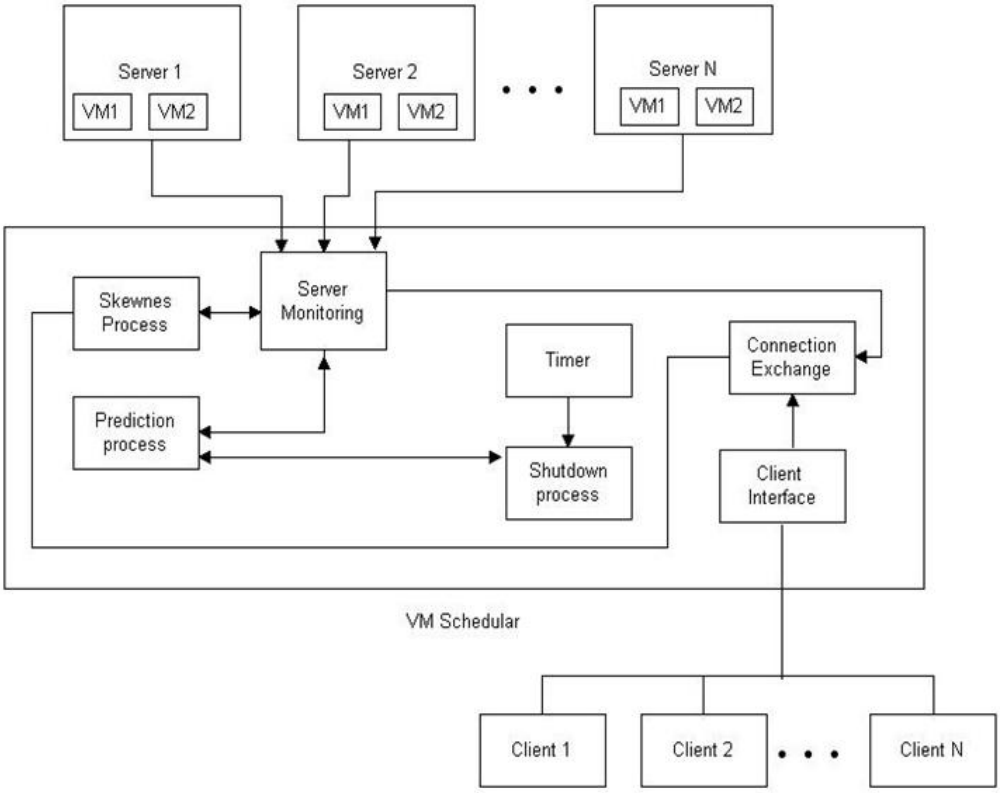
To perform skewness algorithm, we are categorizing the server status spot as:

- Hot spot-: We define a server as a hot spot if the utilization of any of its resources is above a hot threshold. This indicates that the server is overloaded and hence some VMs running on it should be migrated away.
- Cold spot-: We define a server as a cold spot if the utilizations of all its resources are below a cold threshold. This indicates that the server is mostly idle and a potential candidate to turn off to save energy.
- Warm spot-: We define a server as a warm spot if the utilizations of all its resources are below a hot threshold and below cold threshold. This indicates that the server is ready to run VMs.

For each scheduling round, the skewness takes two steps 1) Hot Spot Mitigation 2) Greencomputing to calculate a migration list.
In hot spot mitigation, we first migrate the process from the server whose temperature is hottest or which resource utilization is more. for each hot spot we try to migrate the away VM that can reduce the servers

temperature the most.We choose the server with most skewness reduction. Hot spot mitigation is finished after all hot spots are resolved.

Next the green computing step is invoked here we will try to solve the cold spots in ascending order of memory utilization. To resolve cold spot all its VMs have to be migrated such that the resource utilization should be below the warm threshold after accepting the VM. We also restrict the number of cold spots that can eliminated in each run of the algorithm



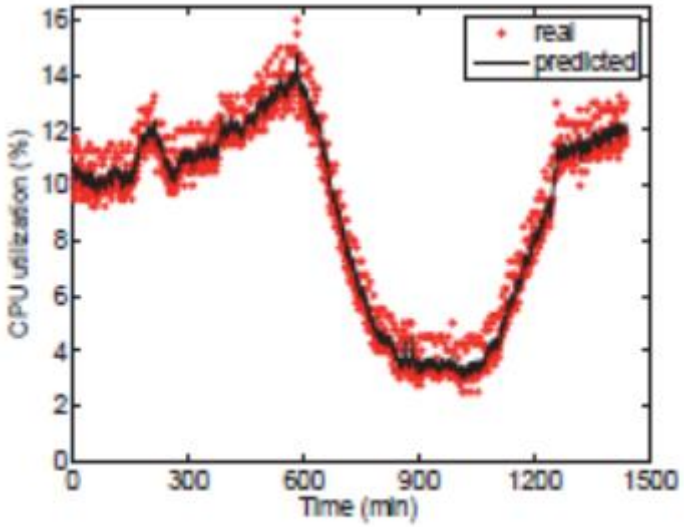
V. LOAD PREDICTION

The prediction algorithm is used to predict the future needs of VMs since we can't shut down or migrate a process to other server. One solution is to look inside the VM for application level statistics. Here we are making our prediction based on the past external behaviours of VMs. In this project, the exponentially weighted moving average (EWMA) prediction algorithm plays an important role in improving the stability and performance of our resource allocation decisions. Based on Physical machines (PMs) Usage we will select server using EWMA algorithm.

The Estimated load and Observed load values are founded by measuring the load every minute and predict the load in the next minute. The load prediction is based on three errors. Median error is calculated on percentage of observed values. Higher and Lower errors are calculated based on predicted values.

TABLE 2
Parameters in our simulation

symbol	meaning	value
h	hot threshold	0.9
c	cold threshold	0.25
w	warm threshold	0.65
g	green computing threshold	0.4
l	consolidation limit	0.05



(a) EWMA: $\alpha = 0.7, W = 1$

Each dot in figure is an observed value and the curve represents the predicted values.The Curve cuts through the middle of the dots which indicates the accurate prediction.

VI. CONCLUSION

We have employed the resource provisioning in cloud computing in which we have reached the goal of achieving the sustainability and green computing in our system.We have also used the skewness concept to combine the VM's so that all the servers are utilized. And Prediction algorithm to predict the future demands.

VII. REFERENCES

[1]. C. Clark, K. Fraser, S. Hand, J. G. Hansen, E.Jul, C. Limpach, I. Pratt, and A. Warfield, "Live migration of virtual machines," in Proc. of the Symposium on Networked Systems Design and Implementation (NSDI'05), May 2005.

[2]. C. A. Waldspurger, "Memory resource management in VMware ESX server," in Proc. of the symposium on Operating systems design and implementation (OSDI'02), Aug. 2002.

[3]. G. Chen, H. Wenbo, J. Liu, S. Nath, L. Rigas, L. Xiao, and F. Zhao, "Energy-aware server provisioning and load dispatching for connection-intensive internet services," in Proc. of the USENIX Symposium on Networked Systems Design and Implementation (NSDI'08), Apr. 2008

[4]. N. Bobroff, A. Kochut, and K. Beaty, "Dynamic placement of virtual machines for managing sla violations," in Proc. of the IFIP/IEEE International Symposium on Integrated Network Management (IM'07), 2007.

[5]. J. S. Chase, D. C. Anderson, P. N. Thakar, A. M. Vahdat, and R. P. Doyle, "Managing energy and server resources in hosting centers," in Proc. of the ACM Symposium on Operating System Principles (SOSP'01), Oct. 2001.

[6]. T. Das, P. Padala, V. N. Padmanabhan, R. Ramjee, and K. G. Shin, "Litegreen: saving energy in networked desktops using virtualization," in Proc. of the USENIX Annual Technical Conference, 2010.

[7]. Y. Agarwal, S. Savage, and R. Gupta, "Sleepserver: a software-only approach for reducing the energy consumption of pcs within enterprise environments," in Proc. of the USENIX Annual Technical Conference, 2010.

[8]. A. Singh, M. Korupolu, and D. Mohapatra, "Server-storage virtualization: integration and load balancing in data centers," in Proc. of the ACM/IEEE conference on Supercomputing, 2008.

[9]. T. Sandholm and K. Lai, "Mapreduce optimization using regulated dynamic prioritization," in Proc. of the international joint conference on Measurement and modeling of computer systems (SIGMETRICS'09), 2009.

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Under Water Image Dehazing Using Gaussian Filter and Laplacian Pyramid

Ammu Bhuvana D, Dr Heena Kousar, Madhu Shree R
Department of CSE, EPCET, Bangalore, Karnataka, India

ABSTRACT
This paper proposes a novel single-image approach for enhancing underwater images that suffer from degradation due to medium scattering and absorption. The method does not require specialized hardware or knowledge of underwater conditions or scene structure. It relies on blending two images that are derived from a color-compensated and white-balanced version of the original degraded image. Weight maps are used to promote the transfer of edges and color contrast to the output image, and a multiscale fusion strategy is employed to avoid artifacts in the reconstructed image. Qualitative and quantitative evaluations show that the enhanced images and videos exhibit better exposedness of dark regions, improved global contrast, and edge sharpness. Furthermore, the algorithm is reasonably independent of camera settings and improves the accuracy of several image processing applications, such as segmentation.
Keywords: under water dehazing, image enhancement, color correction, fusion

I. INTRODUCTION

The underwater environment is important for several reasons. It is home to a diverse range of marine life, which plays a crucial role in the ecological balance of our planet. Underwater ecosystems also provide important resources for human beings, such as food, medicine, and minerals. In addition, underwater environments offer many unique attractions for tourism and recreational activities, such as scuba diving and snorkeling.

In addition, underwater research is important for various scientific fields, including marine biology, oceanography, and environmental science. Studying the underwater environment can help us better understand our planet, its history, and its future. It can also help us identify and address environmental problems, such as pollution and climate change, that affect both marine and human life. Finally, the development of technologies and techniques for exploring and working in the underwater environment can have many practical applications, such as offshore oil and gas exploration, underwater construction, and defense operations.

Underwater imaging faces unique challenges due to light attenuation caused by absorption and scattering effects. This results in poor visibility, with distant objects appearing misty and colors appearing faded. Traditional enhancement techniques are limited, and additional challenges such as color casts and backscatter

further reduce visibility. Underwater imaging is also complicated by movement, water currents, and specialized equipment and techniques are required to capture clear and detailed images.

This paper introduces a novel approach for removing haze in underwater images using a single, conventionally captured image. The approach involves fusing two inputs: a contrast-corrected version and a sharpened, white-balanced version. The white balancing stage removes color distortion caused by light scattering. The multi-scale fusion process ensures a seamless blending without artifacts, resulting in a more natural-looking image. This method offers an effective solution for enhancing underwater image quality.

II. METHODS AND MATERIAL

A. Color Correction

Color correction is a method used to correct the color distortion in underwater images caused by the scattering of light in water. This distortion can lead to images that appear bluish-green or yellowish. Color correction aims to adjust the color balance of the image to make it appear more natural and accurate.

There are two types of color correction methods: statistical analysis and histogram equalization. Statistical analysis methods use statistical techniques to analyze the color distribution of the image and adjust the color balance accordingly. Histogram equalization methods adjust the histogram of the image to improve the contrast and visibility of the image.

B. Fusion

Fusion is a method used to combine multiple images of the same scene to produce a single image that has improved visibility and clarity. This method can be particularly useful in underwater imaging, where the images can be degraded due to the scattering of light in water. There are two types of fusion methods: image-based fusion and sensor-based fusion. Image-based fusion methods combine multiple images of the same scene that have different exposure settings or perspectives. Sensor-based fusion methods combine optical and acoustic sensors. Fusion methods can be effective in improving the visibility and sharpness of underwater images.

III. LITERATURE SURVEY

Simon Emberton et.al [1] This paper presents a new method for improving the visibility in underwater images and videos by identifying and separating the regions that contain only water, known as pure haze regions. These regions have a color that is similar to the haze that needs to be removed for dehazing. The paper proposes a semantic white balancing approach that uses the dominant color of the water to address the spectral distortion in underwater scenes. To validate the effectiveness of the proposed method, the researchers conducted extensive subjective evaluation tests using images and videos captured in different water types.

Y. Liu, S. Rong, et al[2] The quality of underwater images is often poor due to low visibility caused by absorption and scattering, which creates haze and other limitations. To address these issues, a novel dehaze method is proposed in this paper. The method is based on the observation that most pixels in

underwater images tend to distribute nearby a specific plane in RGB space. This observation is called the color space dimensionality reduction prior.

By projecting all pixels to this plane, known as the UV color space, the color distribution can be reduced from the three-dimensional space (RGB space) to a new two-dimensional space (UV space) without causing excessive color shift. By carefully setting the haze-free boundary in UV space, the image transmission can be obtained and an excellent dehazed image can be produced. The results of experiments show that this method is competitive compared to other mainstream underwater single image dehazing methods.

J. Y. Chiang and Y. -C. Chen[3] This paper proposes a new approach to fixing these problems usingg a dehazing algorithm. It estimates the distance between objects and the camera, segments the foreground and background of the image, and checks for the presence of artificial lighting. After compensating for artificial light, it corrects the haze phenomenon and color changes along the underwater propagation path. It then estimates the water depth in the image scene and adjusts color based on the amount of attenuation of each light wavelength.

IV. PROPOSED SYSTEM

Step 1: Input image is taken. Here there are various types of images that are collected and used. The data sets used here are obtained from UIEB, TURBID, and WHOI(Woods Hole Oceanographic Institution).

Step 2: The input image is pre-processed using histogram equalization and contrast stretching. Then, dehazing is performed on the second input image using the dehaze function where the input image is subjected to contrast stretching and then sent to guided filter for haze removal. Here, the haze removal process is fastened by the help of box filter. The first and second input images are then displayed.

Step 3: A weight map is generated for both the histogram equalized image and the contrast stretched image. The luminance and saliency weight maps are calculated for the first (Histogram equalized image) and the second (contrast stretched image after haze removal) input image. The resultant weight maps for both inputs are calculated by adding the luminance and saliency weight maps.

Step 4: The normalized weight maps for both inputs are calculated by dividing the resultant weight maps by the sum of the resultant weight maps. Gaussian pyramids are then generated for the normalized weight maps.

Step 5: Laplacian pyramid decomposition is performed on each color channel of both input images separately. The Laplacian pyramid consists of several levels of high-pass filtered images, each level capturing different scales of detail in the image.

Step 6: The fusion process is then performed by multiplying each level of the Laplacian pyramid of each color channel with the corresponding level of the Gaussian pyramid of the normalized weight map for each input image. The resultant pyramids are then added to obtain the fused pyramid.

Step 7: A fused image is reconstructed by performing pyramid reconstruction on each color channel separately. The reconstructed color channels are then combined to obtain the final fused image.

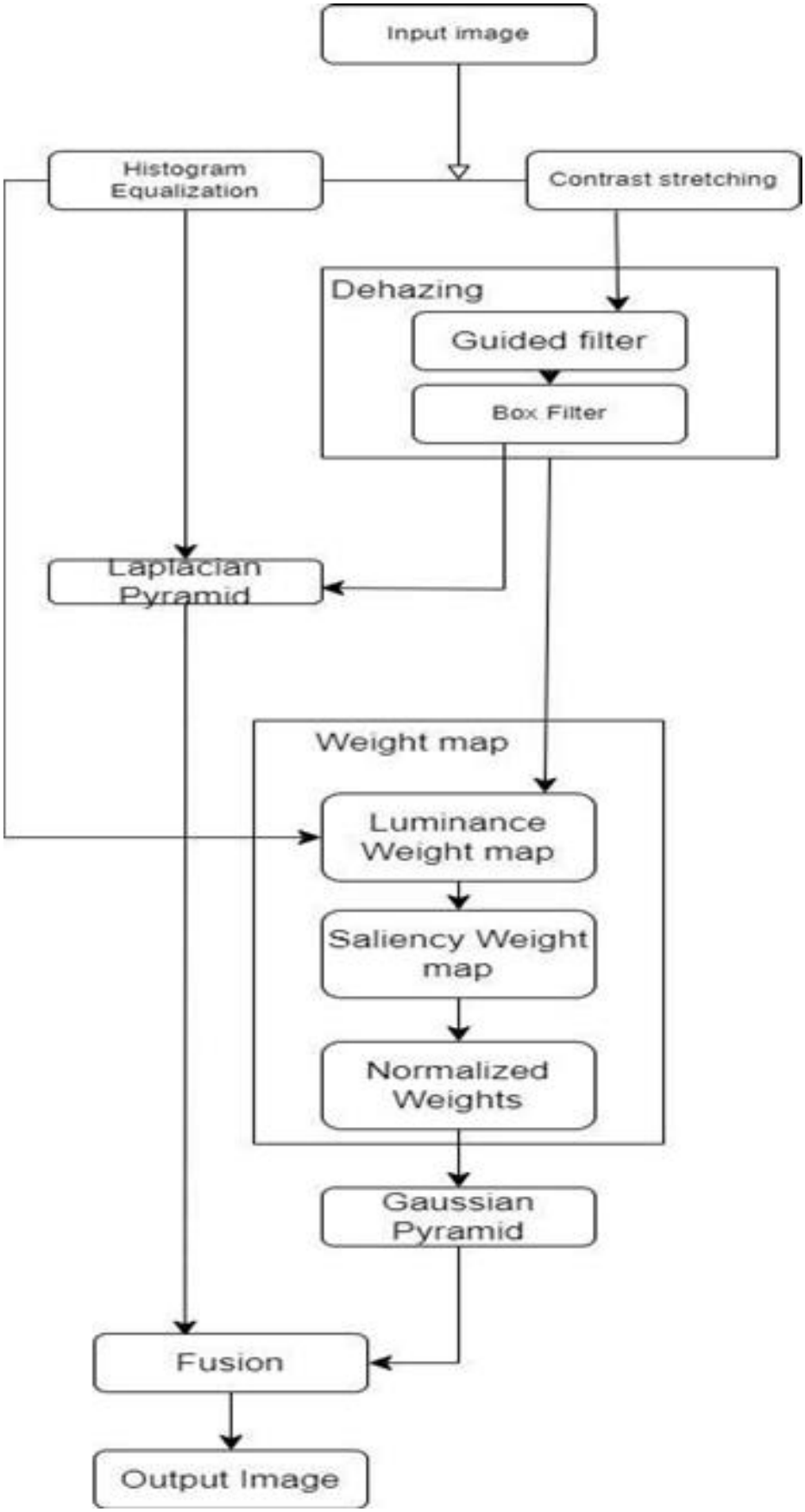
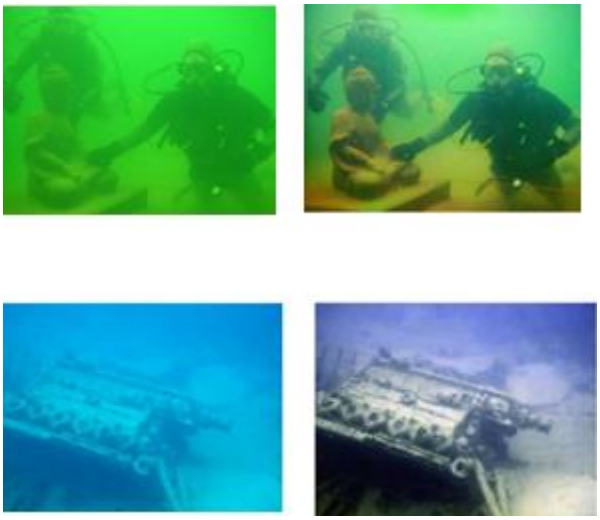


Figure 1:Proposed system

V. RESULTS



Original image Dehazed image

VI. CONCLUSION

It enhances underwater images to correct color distortion caused by attenuated wavelengths, fog, blur and other factors. The image is clearer and can be used by researchers to study about marine world

VII. REFERENCES

[1]. C.O.Ancuti, C. Ancuti, C. D. Vleeschouwer and P. Bekaert, Color balance and fusion for underwater image enhancement, IEEE Transactions on Image Processing 27 (Jan 2018).

[2]. S. Bazeille, I. Quidu, L. Jaulin and J. P. Malkasse, Automatic underwater image pre-processing, Proceedings of 1900(1) (2006)

[3]. G. Buchsbaum, A spatial processor model for object colour perception, Journal of the Franklin Institute 310(1) (1980)

[4]. P. J. Burt and E. H. Adelson, The laplacian pyramid as a compact image code, in Readings in Computer Vision (Elsevier, 1987)

[5]. N. Carlevaris-Bianco, A. Mohan and R. M. Eustice, Initial results in underwater single image dehazing, MTS/IEEE Seattle, OCEANS 2010 (2010).

[6]. J. Y. Chiang and Y. C. Chen, Underwater image enhancement by wavelength compensation and dehazing, IEEE Transactions on Image Processing (2012)

[7]. L. Chao and M. Wang, Removal of water scattering, ICCET 2010 - 2010 International Conference on Computer Engineering and Technology, Proceedings 2 (2010)

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Preserving Privacy in EHR Sharing with Consortium Blockchain and Searchable Encryption

Heena Kousar*, Ammu Bhuvana S, Madhushree, Khallikkunaisa
Department of Computer Science & Engineering, VTU/EPCET/Bengaluru, Karnataka, India

ABSTRACT
This research paper presents an innovative protocol that utilizes blockchain technology to facilitate secure and privacy-preserving sharing of Electronic Health Records (EHRs). By harnessing the decentralized, anonymous, and verifiable nature of blockchain, the protocol effectively tackles the data security and privacy concerns commonly associated with cloud-based EHR sharing systems. The protocol offers data requesters the capability to search for relevant EHRs on the EHR consortium blockchain using specific keywords. With explicit authorization from the data owner, the requesters can retrieve re-encryption ciphertext from the cloud server. To ensure robust data security, privacy preservation, and access control, the protocol incorporates searchable encryption and conditional proxy re-encryption techniques. To guarantee system availability, a proof of authorization mechanism is implemented as the consensus mechanism for the consortium blockchain. This mechanism rigorously verifies and validates the authorization of data requesters, thus enhancing the overall security and integrity of the system.
Keywords: Datasharing, Electronic Health Records, Privacy preservation, Blockchain

I. INTRODUCTION

EHR sharing has attracted significant attention and research from industry and academia, particularly in the areas of privacy preservation, data security and interoperability [2]. Privacy preservation is crucial as EHRs contain personal and highly sensitive information that must be protected to safeguard patients reputation and well-being. Similarly, data security is essential to ensure that only authentic information is included in EHRs, as forged or modified data can undermine the effective utilization of these records. Lastly, interoperability plays a vital role in allowing patients to have control over the access rights of their EHRs and facilitating the seamless exchange of records between different healthcare institutions.

A HER sharing protocol based on a consortium blockchain is proposed which ensures secure and privacy-preserving data access. Only authorized data requesters who possess the search trapdoor can obtain relevant keywords and information. The blockchain accounts handle authorization and other access services, guaranteeing privacy protection. Additionally, the cloud re-encrypts the EHR ciphertext and securely shares it with the agreed-upon data requester.

In this paper, we first provide an overview of related research in Section II, Section III presents the key technologies necessary for our protocol. Section IV discusses the system architecture, EHR consortium blockchain, threat model and security goals. It is then delve into the protocol details and security proof in Section V. Atlast summarize the paper and discuss the future implementation.

II. RELATEDWORK

In this section, we discuss works that focus on EHR sharing with the help of cloud technology and blockchain technology

A. EHR SHARING WITH CLOUD

Access control schemes and searchability and interoperability of EHR sharing have been improved using cloud technology and blockchain technology. To ensure data security during the process of EHR sharing, access control schemes based on the cloud have been introduced [3]-[5]. A study proposed a new method called ciphertext-policy attribute-based sign-encryption and secure sharing of personal health records in cloud computing, which enables fine-grained access control. Another study [4] proposed an efficient and secure fine-grained access control scheme for authorized users to access EHRs in cloud storage. Furthermore, [5] developed a hierarchical comparison-based encryption scheme and a dynamic policy updating scheme using the proxy-encryption technique to achieve dynamic access control in cloud-based EHR systems.To enhance the searchability and interoperability of EHR [6] sharing, a study presented a cloud-based EHR system that supports fuzzy keyword search, ensuring secure data sharing and effective utilization of the EHRs. Another study utilized conjunctive keyword search with proxy re-encryption to create a secure EHR search scheme for data sharing between different medical institutions. Moreover, proposed a general framework for secure sharing of EHRs, allowing patients to securely store and share their EHRs on a cloud server, while enabling doctors to access the EHRs in the cloud. In addition, a study proposed a blockchain-based secure and privacy-preserving EHR sharing protocol, which allows data requester to search desired keyword from EHRs without revealing the content of the EHRs. The protocol ensures data security and privacy preservation via consortium blockchain.

B. EHR SHARING WITH BLOCKCHAIN

Data Security: Blockchain technology can enhance the security of electronic health records (EHRs) by providing a tamper-resistant and immutable ledger. Each transaction or modification made to the EHR is recorded as a block, linked to previous blocks, and distributed across a network of computers. This decentralized nature makes it difficult for malicious actors to alter or manipulate the data without consensus from the network [10]. Interoperability: Blockchain has the potential to improve interoperability between different healthcare systems. Since blockchain operates on a distributed ledger, it can facilitate seamless data exchange and integration among various healthcare providers, regardless of their underlying EHR systems. This can help eliminate data silos and improve care coordination.

Data Integrity: Blockchain's immutability ensures the integrity of EHRs. Once data is recorded on the blockchain, it becomes virtually impossible to modify or delete it without leaving a trace. This feature can help address issues related to data tampering and ensure the accuracy and reliability of patient health records.The

architecture mentioned, which combines blockchain technology with intelligent contracts and user-generated acceptable policies, can be a promising approach to enhancing the security control of personal data in health information exchange. By leveraging blockchain's decentralized and immutable nature, along with smart contracts' automation capabilities, it is possible to create a robust framework for data security [11]. The conceptual design you mentioned, which combines blockchain technology with cloud storage, aims to facilitate safe and transparent sharing of personal continuous-dynamic health data. Here's a high-level overview of this design:

Blockchain Network: The design utilizes a blockchain network as a decentralized and immutable ledger. Health data is encrypted and stored on the blockchain, ensuring its integrity and protection from unauthorized access. The blockchain's transparency allows for auditing and traceability of data transactions.

Cloud Storage: In addition to the blockchain, the design incorporates cloud storage for efficient and scalable data storage. While the blockchain stores metadata and references to health data, the actual data itself is securely stored in the cloud. Cloud storage enables the storage of large volumes of continuous and dynamic health data.

Secure Data Sharing: Users can securely share their personal health data through the blockchain network. The blockchain facilitates the creation of secure and transparent transactions when sharing data. Users can define access permissions and consent requirements using smart contracts, ensuring that data is shared only with authorized parties and with the necessary consent.

III. PRELIMINARIES

The technical preliminaries mentioned in this section pertain to the concepts of bilinear maps and complexity assumptions. Here's a breakdown of the key definitions and properties:

Bilinear Maps:

Bilinear Map (e^): A bilinear map is denoted as e^ and operates between two cyclic groups of the same prime order. It takes two elements from the first group (F1) and maps them to an element in the second group (F2).

Admissible Bilinear Map: An admissible bilinear map, denoted as e^, satisfies the following properties: a. e^(aR, bS) = e^(R, S)^(ab) for all R, S ∈ F1 and a, b ∈ Zq* (non-zero integers modulo q). b. e^(R, S) = e^(S, R) (symmetric property). c. e^(R + S, T) = e^(R, T) * e^(S, T) for all R, S, T ∈ F1 (bilinearity property). d. There exist elements R, S ∈ F1 such that e^(R, S) ≠ 1 (mod F2). e. The bilinear map e^ can be efficiently computed.

Complexity Assumptions:

Elliptic Curve Discrete Logarithm Problem (ECDLP): In this problem, we consider an elliptic curve denoted as E. The primitive element on the curve is P, and X is another element on the elliptic curve. Given pE as the number of points on the curve, the ECDLP aims to find an integer b (1 ≤ b ≤ pE) such that P + P + ... + P = bP = X. The ECDLP is assumed to be computationally difficult, meaning that finding the value of b from the given elements is challenging and no efficient algorithm exists for solving it.

In cryptosystems, the private key is usually an integer, and the public key is a point on the curve with coordinates X (xX,*yx*).

ECDLP Assumption. It is assumed that it is difficult to solve the ECDLP in polynomial time.

Definition 2: The Decision Linear Diffie-Hellman Problem (DLDH) is defined in the context of an elliptic curve E and a cyclic group G1 of prime order Q. Let P1, P2, and P3 be random elements in G1, and let a1, a2, and a3 be random numbers in Zq*. The DLDH problem is formulated as follows:

Given the tuple (P1, P2, P3, a1P1, a2P2, a3P3) ∈ F1 as input, the goal is to determine whether a3 is equal to a1 + a2 or not. The output is 1 if a3 = a1 + a2, and 0 otherwise.

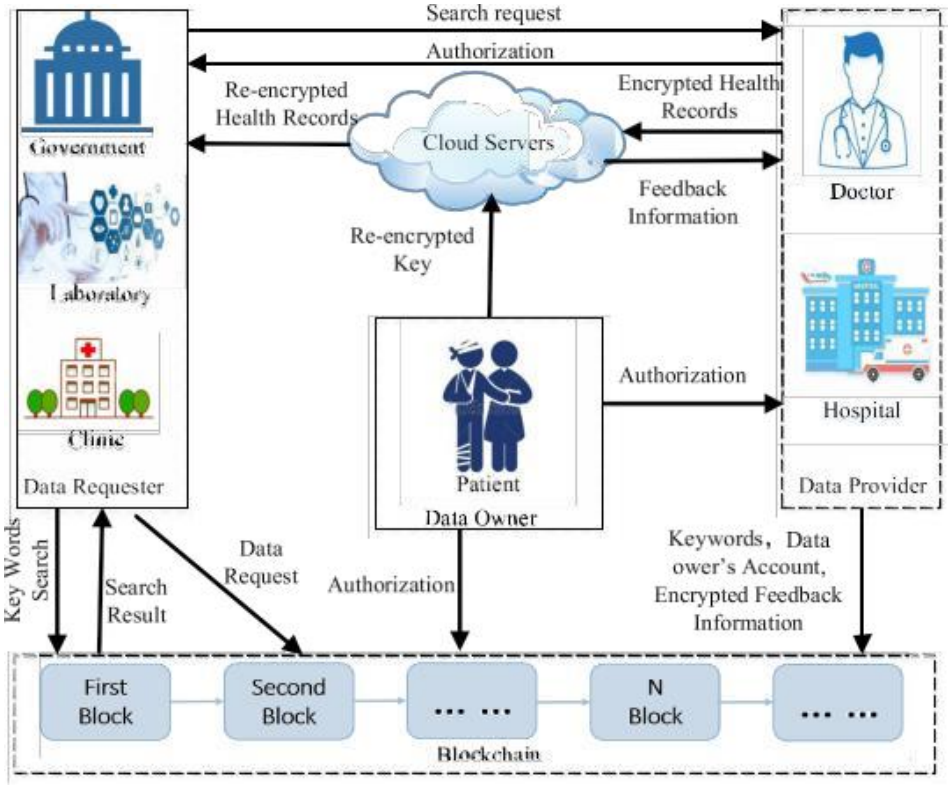
The advantage of an algorithm A in deciding the DLDH problem in F1 is defined as the probability that A correctly determines the equality a3 = a1 + a2 when given the input tuple (P1, P2, P3, a1P1, a2P2, (a1 + a2)P3). Mathematically, it is expressed as:

Pr[A(P1, P2, P3, a1P1, a2P2, (a1 + a2)P3) = 1]

The advantage of an algorithm A represents how well it can solve the DLDH problem, with a higher advantage indicating a better ability to distinguish valid solutions from invalid ones.

IV. SYSTEM ARCHITECTURE

There are five entities in the proposed framework: Data owners (DO), data providers (DP), cloud servers (CS), blockchain (BC), and data requesters (DR), as shown in Fig. 1.



- **DATAOWNERS**

In the given context, "DO" refers to data owners, who are the patients visiting doctors in hospitals or medical institutions for medical services. The data owners are the source of health records, and as such, they have ownership and control rights over their data. To facilitate data sharing, the data owners need to register an account on the EHR (Electronic Health Records) consortium blockchain.

- DATA PROVIDERS

In the given context, "DO" refers to data owners, who are the patients visiting doctors in hospitals or medical institutions for medical services. The data owners are the source of health records, and as such, they have ownership and control rights over their data. To facilitate data sharing, the data owners need to register an account on the EHR (Electronic Health Records) consortium blockchain.

If a new DP wants to join the blockchain, he or she has to take three steps:

- Register an account with the EHR consortium.•
- Submit a recommendation letter signed by one commissioner and send it to all of the•
- Get at least 2⁄3 of the authorizations from commissioners.

- CLOUD SERVER

The cloud server's primary role in the system is to securely store the encrypted Electronic Health Records (EHRs) provided by the Data Providers (DPs). It ensures that the stored files are kept confidential and intact, and it also shares the file location information with the Data Owners' (DOs) accounts on the EHR consortium blockchain. However, it's worth noting that although the cloud server operates honestly and follows the prescribed procedures, it may have a natural curiosity about the data it handles. Additionally, the cloud server is responsible for re-encrypting the EHRs using the appropriate re-encryption key when required, further safeguarding the privacy and security of the records.

- DATA REQUESTERS

Data requesters, which can include government entities, laboratories, clinics, and other relevant organizations, are key stakeholders in the proposed system for accessing patients Electronic Health Records (EHRs). Their role involves obtaining a search trapdoor from the Data Provider (DP) to search for specific keywords within the blockchain. Based on the search results, they can then send a formal request to the Data Owner (DO) for authorization to access the desired EHRs.

Once the data requesters receive authorization from the DO, they are granted access to the encrypted health records stored on the cloud server. These health records are securely maintained to ensure confidentiality and integrity. As part of their interaction with the system, the data requesters' actions generate service transactions, which are added to the transaction pool. This active involvement positions them as service transaction senders within the blockchain network.

Data requesters have the flexibility to join or exit the blockchain network at any time, functioning as regular users. They have visibility into the entire consensus process and can observe the system's operation. Additionally, they have the privilege of enjoying the services provided by the system

V. PROPOSED MODEL

The process of the proposed protocol is represented in Fig. 2.The protocol is made up of three layers: Data generation layer, data storage layer, and data sharing layer.

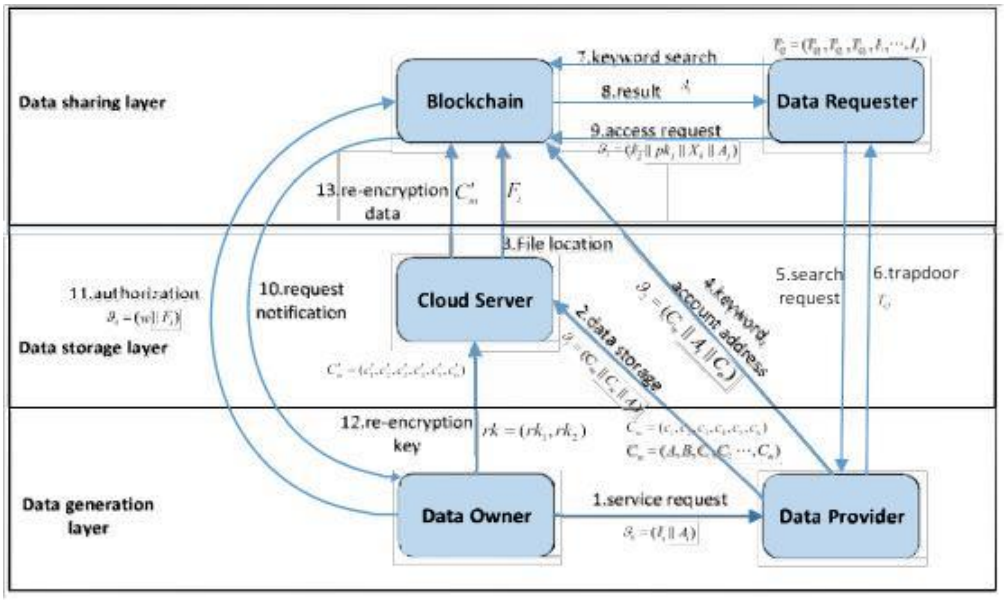


Fig 2: Proposed Protocol

When a patient, referred to as DO with identity Ii, visits a hospital for medical services, they are required to register an account in the EHR consortium blockchain. The EHR consortium blockchain generates an account address and private key, which are then sent to the patient for secure access.

The patient, denoted as i, transmits a data packet ϑ0 = (Ii || Ai) to a doctor, identified as k. The original Electronic Health Record (EHR) for patient i is generated through interactions with doctor k, who is the Data Provider (DP). The DP extracts a series of keywords from the EHR. Subsequently, the DP encrypts the EHR, denoted as m, using the patient's public key (pki), the DP's private key (xk), and the keyword wi. This encryption process produces the EHR ciphertext (Cm). Additionally, the DP encrypts the keyword wi using their own public key (Xk), resulting in the keyword ciphertext (Cw). The DP then sends the data packet ϑ1 = (Cm Cw Ai) to the cloud server for storage. Once the cloud server safely stores the data, it sends the file location (Fi) to the DO's account.

When data requesters (DRs) need to search for specific Electronic Health Records (EHRs), they begin by submitting a search request to the Data Provider (DP). If the request is approved, the DRs are provided with a trapdoor (TQ) to facilitate their search. Using this trapdoor, the DRs can explore the blockchain and locate the matched EHRs, along with the associated Data Owner's (DO) account address (Ai).

Afterwards, the DRs initiate an access request by sending a data packet ϑ3 = (Ij pkj Xk Aj) to the DO's account. Upon receiving this request, the DO responds by granting authorization. The authorization includes essential details such as the file location (Fi) and the corresponding keyword (wi). Additionally, the DO generates a re-encryption key (rk) and shares it with the Cloud Server (CS). The CS utilizes this key to perform proxy re-encryption on the required ciphertext.

Lastly, the DRs utilize their private key (skj) to decrypt the re-encrypted ciphertext (Cmr), enabling them to securely access the desired EHRs and retrieve the relevant information.

VI. CONCLUSION

In our research, we have proposed an original scheme for sharing Electronic Health Records (EHRs) using blockchain technology. Our approach focuses on ensuring data security and privacy preservation during the sharing process across diverse medical institutions. We have introduced a framework that combines cloud-assisted storage and blockchain to facilitate efficient EHR sharing between entities. The cloud serves as the storage provider for the encrypted EHR ciphertext, while the EHR indexes are stored securely on the EHR consortium blockchain. To address data security and privacy concerns, we have designed an architecture and protocol that incorporate conjunctive keyword-searchable encryption and conditional proxy re- encryption. These mechanisms enable efficient searchability of EHRs based on multiple keywords while ensuring secure data transfer between authorized parties.

For future work, our plan is to implement and optimize the proposed scheme using the Hyperledger Fabric blockchain platform. This will involve refining the smart contracts responsible for executing the data sharing algorithms, thereby enhancing the functionality and performance of our system.

should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence. We suggest that you use

Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions. Authors are strongly encouraged not to call out multiple figures or tables in the conclusion these should be referenced in the body of the paper.

VII. REFERENCES

[1]. H. K. Patil and R. Seshadri, "Big data security and privacy issues in healthcare," in Proc. IEEE Int. Congr. Big Data, Anchorage, AK, USA, Jun./Jul. 2014, pp. 762–765.

[2]. J. Li and X. Li, "Privacy preserving data analysisin mental health researc," n Proc. IEEE Int. Congr. Big Data, New York, NY, USA, Jun./Jul. 2015 pp. 95–101.

[3]. J. Liu, X. Huang, and J. K. Liu, "Secure sharing of Personal Health Records in cloud computing: Ciphertext-policy attribute-based signcryption," Future Gener. Comput. Syst., vol. 52, pp. 67–76, Nov. 2015.

[4]. X. Liu, Y. Xia, W. Yang, and F. L. Yang, "Secure and effificient querying over personal health records in cloud computing," Neuro Comput., vol. 274, pp. 99–105, Jan. 2018.

[5]. X. Liu, Q. Liu, T. Peng, and J. Wu, "Dynamic access policy in cloud based personal health record (PHR) systems," Inf. Sci., vol. 379, pp. 62–81, Feb. 2017.

[6]. Z. Liu, J. Weng, J. Li, J. Yang, C. Fu, and C. F. Jia, "Cloud-based electronic health record system supporting fuzzy keyword search," Soft Comput., vol. 20, pp. 3243–3255, Aug. 2016 .

[7]. X. Wang, A. Zhang, X. Ye, and X. Xie, "Secure-aware and privacy preserving electronic health record searching in cloud environment," Int. J. Commun. syst., vol. 32, p. e3925, May 2019. doi: 10.1002/dac.3925.

[8]. M. H. Au, T. H. Yuen, J. K. Liu, W. Susilo, X. Huang, Y. Xiang, and Z. L. Jiang, "A general framework for secure sharing of personal health records in cloud system," J. Comput. Syst. Sci., vol. 90, pp. 46–62, Dec. 2017.

[9]. G. Zyskind, O. Nathan, and A. S. Pentland, 'Decentralizing privacy: Using blockchain to protect personal data," in Proc. IEEE Secur.Privacy Workshops, vol. 90, May 2015, pp. 180–184.

[10]. S. Amofa, E. B. Sifah, K. O.-B. Agyekum, S. Abla, Q. Xia, J. C. Gee, and J. B. Gao, "A blockchain-based architecture framework for secure sharing of personal health data," in Proc. IEEE 20th Int. Conf. e-Health Netw., Appl. Services (Healthcom). Ostrava, Czech Republic, 2018, pp. 1–6.

[11]. X. Zheng, R. R. Mukkamala, R. Vatrapu, and J. Ordieres-Mere, "Blockchain-based personal health data sharing system using cloud storage," in Proc. IEEE 20th Int. Conf. e-Health Netw., Appl. Services (Health com). Ostrava, Czech Republic, Sep. 2018, pp. 1–6.

[12]. T. Mikula and R. H. Jacobsen, "Identity and access management with blockchain in electronic healthcare records," in Proc. 21st Euromicro Conf. Digit. Syst. Design (DSD), Prague, Czech Republic, 2018, pp. 699–706.

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Hadoop MapReduce Framework for Performance Aware Scheduling

Jagadevi Bakka*, Sanjeev C Lingareddy, Manimegalai A, Rashmi T V
Department of CSE, East Point College of Engineering and technology, Bangalore, Karnataka, India

ABSTRACT
MapReduce (MR) has been one of the popular computing framework for BigData analysis and processing application in last decade; further Hadoop is an open source platform which is widely used for MR framework. Moreover, existing HMR aka Hadoop-MR model faces major issues like I/O overhead and memory overhead. In this research work, we focus on developing memory and performance awarescheduler design named as MPA-HMR for efficient utilization of system resources and data processing in real time. MPAS-HMR is developed for analyzing the Global Memory Management; thus minimizing the Disk I/O seek. Moreover, MPAS method are evaluated on the Microsoft Azure HDInsight cloud platform in consideration with text mining applications, also comparative analysis with the existing model is carried out. Further, comparative analysis shows that our model outperforms existing model in terms of computation time and computing cost.
Keywords: Cloud computing, MapReduce, Performance modelling, Resource utilization, Task scheduling.

I. INTRODUCTION

Several organizations such as educational institution, government and industry gathers huge data through various sources like WWW, bioinformatics, social network, sensor network and so on for different purpose. Moreover, analyzing these unstructured data has become one of the desired work for various organization; however state-of-art approach fails to perform considering the real time scenario on the stream/continuous data. In case of real time scenario, data-based platform like google have designed the parallel computational approach named MR (MapReduce) framework [Dean and Ghemawat (2008)]; this particular framework offers parallel execution in distributed manner. HMR (Hadoop MapReduce) is one of the popular and widely adopted tool in comparison with other tools like Phoenix [Taura et al. (2003)], Mars [He et al. (2008)] and Dryad [Isard et al. (2007)]; as HMR is open source [Kang et al. (2011)].
HMR model comprises various phases which includes Setup, Mapping, shuffling and reduce; these are shown in Fig. 1; moreover, HMR have computing nodes cluster and master node. Further, Jobs assigned to Hadoop are shared into Mapping and reducing tasks; in setup phase, input data are divided into particular volume known as chunks for Map nodes. Furthermore, Hadoop parts MR (MapReduce) jobs into various task set where each chunk are processed through Map Worker; in general Map phase accepts the input in certain form as (K_1,V_1) key/value and creates further pair of key/value as an output. Shuffle phase starts after Map phase completion

where intermediate key and value pair are gathered from Map Task; sorting is carried out on the intermediate pair of key, value. In general sorting and shuffling are combined in shuffling phase, also reduce phase process the data in accordance with UDF (User Defined Function). At last, reduce phase output is written and stored in HDFS aka Hadoop distributed-FS (File system).

In past few years, Hadoop application has seen enormous growth and performance enhancement has been observed as well [Lin et al. (2012)], [Cui et al. (2013)]; there are various model of Hadoop some of the effective methods are presented in [Khan et al. (2014)] developed starfish model that gathers Hadoop task profile for satisfactory granularity. In [Xu et al. (2017)] developed mechanism named Elasticiser which was based on VM considered (as in starfish model) for resource allocation problem; however, it leads to over-predicted task run time and large overhead while gathering the active task profile (Hadoop task). Further, considering this drawback, [Glushkova et al. (2017), Ehsan et al. (2017), Khan (2016)] utilizes overlapping and non-overlapping phenomena and to predict the task, conventional LR (Linear regression) is adopted. Moreover these methods also predicts the amount of resources for different task with deadline as constraint. In [Wu et al. (2017)] found that slow shuffling is main reason for any degradation in MapReduce and only considerable amount of work has been carried out for shuffle phase speed optimization; hence they presented a new mechanism for balancing the network loads on various cross rack links while sampling and shuffling for application where random processing generates efficient results. However, these schemes were designed for sampling-based application only and they were not convenient for general application where whole data is processed.

In [Yao et al. (2019)] introduces YARN mechanism integrated with resource management for scheduling of jobs and they made a point that fairness and efficiency are that major concern in resource management since resources shared by the various applications. Moreover, current scheduling mechanism in YARN does not provide the optimal resource management, hence this framework omits the dependency among the defined which is one of the major concerns for resource utilization and heterogeneous characteristics in real time scenario. In [Zheng et al. (2018)], it is observed map phase is considered to be CPU sensitive whereas I/O intensive and these phases are performed parallelly.

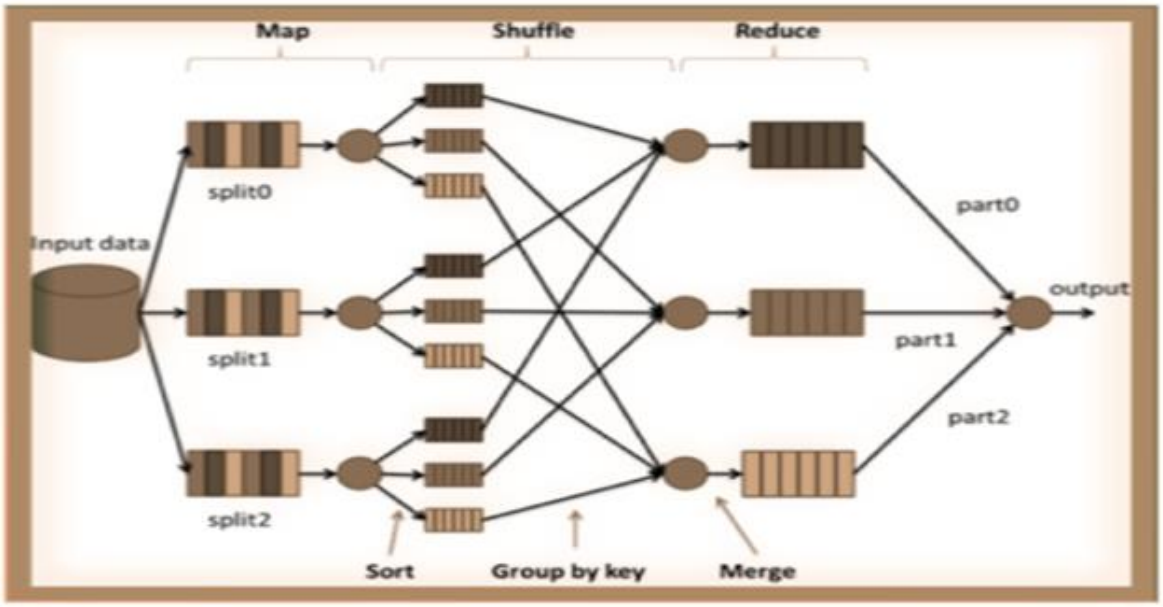


Figure 1. Architecture of Standard Hadoop MapReduce Framework.

Further, author performed joint scheduling for overlapping mapping and shuffling to optimize the makespan. Similarly [Yang et al. (2019)] adopted mechanism which was based on the dynamic scheduling for minimizing the shuffle traffic as several existing methodology failed to consider the impact of data centers. In here, Hit(Hierarchical topology) aware MR (MapReduce) was proposed for reducing the overall traffic cost which in terms reduces the execution time. However, these model do not utilize memory efficiently as jobs are executed at system level rather than thread level.

Moreover to overcome these challenges this research work designs and develops memory constraint aware scheduler for HMR framework namely MPAS-HMR; MPAS-HMR is very much similar to work carried in [Apache (2014)]; further a thread based execution is considered for optimal memory utilization and minimization of I/O overhead [Zhang, J. et al.(2012), Longbin, L., et al.(2013), Kim et al. (2018)], also this research work focuses on developing a dynamic memory distribution among the task throughout thread in one VM. Furthermore, this research work develops I/O model to improvise memory management for CPU and cross I/O, also MPAS-HMR helps in avoiding the re-reading the data before transmission which minimizes the task through caching final outcome of job in memory.

Research Contribution are as follows:

- Presented memory and performance aware scheduling design for HMR for executing text mining and iterative application.
- The proposed MPAS design reduces makespan and computational cost for executing text mining and iterative applications when compared with existing scheduling design for HMR [Yao et al. (2019)].

The rest of the paper is organized as follows. In section II the memory and performance aware scheduling design for HMR framework is presented. In penultimate section experimental study is carried out. The conclusion and future work is described in last section.

II. METHODS AND MATERIAL

In this section, we present a new framework namely, Memory and performance aware scheduler (MPAS) design for Hadoop MapReduce Framework as shown in Figure 2.

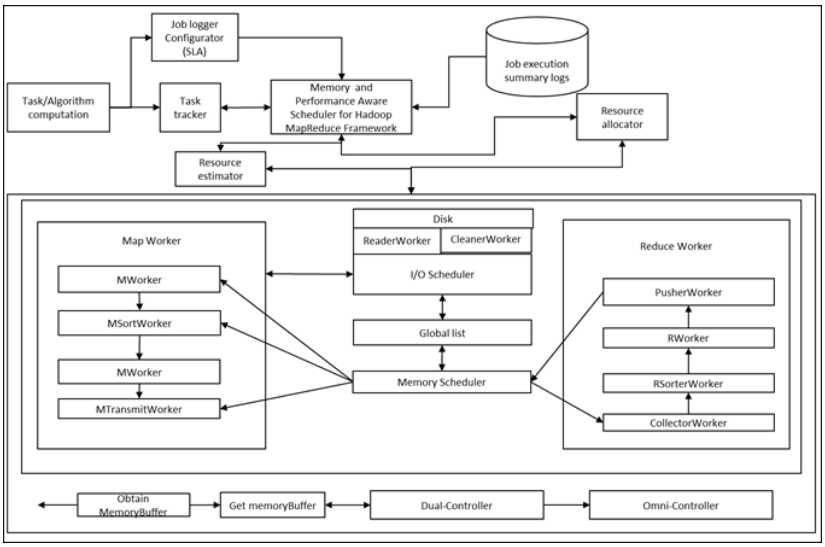


Figure 2. Architecture of Memory and performance Aware Scheduling Design for Hadoop MapReduce Framework.

a) System model:

In standard Hadoop-MR framework, the tasks are executed on different nodes individually. However, in MPAS-HMR framework, the task will be executed through Memoryschedular. The Memoryschedular is responsible for allocation and deallocation of memory resources. Here different worker will have different memory level and these information about memory resource capacity can be collected from GlobalList. The I/O scheduler pings ReadWorker for collecting information from disk and CleanWorker cleans information from the GlobalList. The MPAS-HMR realizes global memory management through GlobalList by adopting such data structure mechanism. In GlobalList, the intermediate data of different task are sorted and kept. For reading and cleaning data from the disk the I/O scheduler uses Multiple-buffers. In this way memory resource are utilized more efficiently aiding in reduction of makespan.

b) Memory and Performance Aware Scheduling design:

The memory and performance aware scheduler is designed considering following assumption. First, buffers size will be of varied size; thus for allocating memory resource to these buffers an effective optimization design must be modelled. Second, different MapReduce tasks will have different memory requirement; thus, dynamic memory allocation design is needed. The total size U_T of different buffers is estimated using CacheList as follows

u_T=T^↑-〖E_(〖list〗_T)-N〗_(D_T) (1)

where T^↑ represent memory size maximal limit for storing intermediate data, E_(〖list〗_T) represent the overall size of DataPairList, and N_(D_T) depicts I/O Scheduler overall memory usage.

In similar manner, the MapController uses memory of size 〖MC〗_T for executing Map task is computed using following equation

〖MC〗_T=min⁡(〖P_D〗_T+〖〖Qtrnsm〗_D〗_T,U_T) (2)

where 〖P_D〗_T represent MSort buffer size and 〖〖Qtrnsm〗_D〗_T defiens I/O buffer size. The MSort buffer size is computed using following equation

〖P_D〗_T={(NP^↑*N_o NP^↑≠0@〖〖QP〗_D〗_T NP^↑=0)┤ (3)

where NP^↑ represent MSort maximal size for executing each task, 〖〖QP〗_D〗_T defines current MSort buffer size and M_n describes the total Map task current being processed. Then, the Reduce Controller memory size 〖RC〗_T for executing task is computed using following equation

〖RC〗_T=T_S-〖MMC〗_S (4)

The MPAS design keep enough memory in reserve for executing task; thus, avoid frequent recycling of memory and I/O resource aiding in reduction of makespan. The makespan C of for executing job can be computed using following equation

C=C_T+C_M+C_R. (5)

where C_T define makespan for initialization worker, C_M depicts map job execution makespan, and C_R define reduce job execution makespan. Let consider that each workerq is composed n number of core/thread with memory size of x; then the average makespan for executing task can be computed using following equation

C_M=(∑_(a=1)^q▒C_(a_M))/q. (6)

Similarly, for reduce task average makespan can be computed as

C_R=(∑_(a=1)^q▒C_(a_R))/q. (7)

Using Eq. (6) and (7), the total makespan of MPAS can be computed as

$C=C_T+(\sum_{a=1}^{q}(C_{a_M}+C_{a_R}))/q. \quad (8)$

The MPAS design minimize makespan and reduce cost for executing text mining and iterative application when compared with existing HMR scheduling methodologies which is experimentally shown below.

III. RESULTS AND DISCUSSION

Here experiment is conducted to evaluate the performance of MPAS-HMR over HaSTE [21]. The system parameter used for experiment analysis is Ubuntu 16 operating system configured with 8GB RAM and two cores. Hadoop cluster with one master and two slave node of identical configuration is used similar to HDInsight Azure A2_v2 instance [24]. Experiment is conducted on simple Wikipedia dataset of size varied from 250 MB to 1 GB. Further, experiment is conducted using complex sensor data of size varied from 100MB to 400 MB. Outcome is measured in terms of make span and computational cost for executing above workload using respective scheduling mechanism.

A. The makespan outcome achieved for executing simple workload of varied size by HaSTE and MPAS-HMR is shown in Fig. 3. MPAS-HMR reduces makespan by 3.714%, 6.66%, and 8.52% when compared with HaSTE when workload size is 100MB, 200MB, and 1000MB, respectively. From result obtained it can be state that MPAS-HMR improves makespan performance by 6.3% on an average when compared with HaSTE.

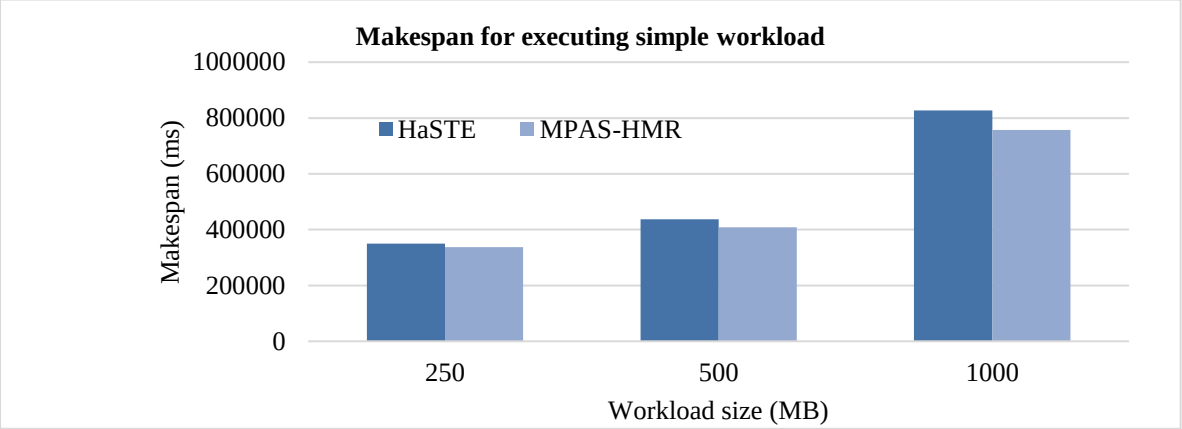


Figure 3. Makespan performance for executing simple workload.

The computational cost induced for executing simple workload of varied size by HaSTE and MPAS-HMR is shown in Fig. 4. MPAS-HMR reduces computational cost by 4.079%, 7.015%, and 8.87% when compared with HaSTE when workload size is 100MB, 200MB, and 1000MB, respectively. From result obtained it can be state that MPAS-HMR reduce computation cost by 6.654% on an average when compared with HaSTE under varied workload scenarios.

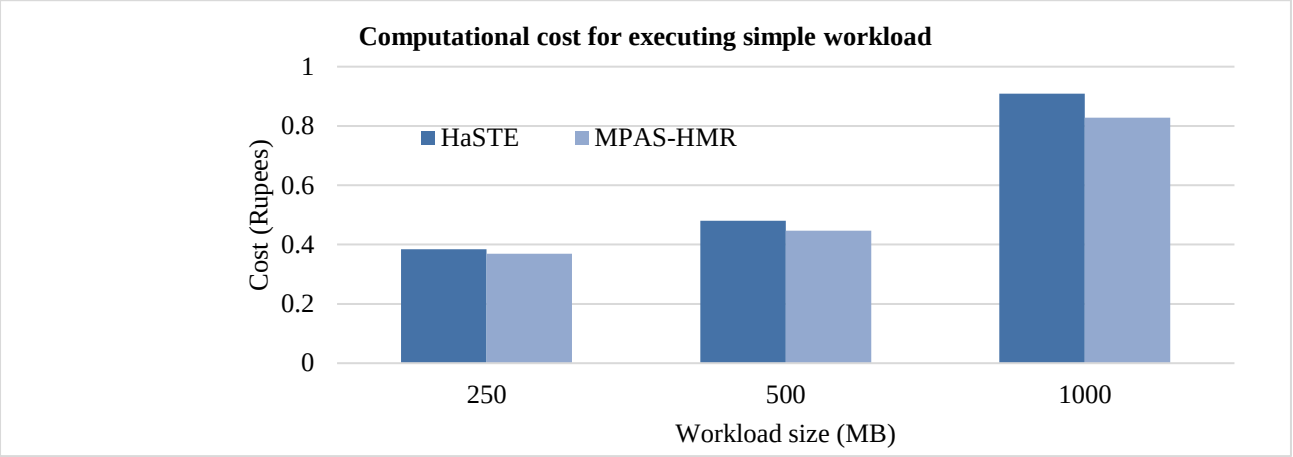


Figure 4. Computational cost for executing simple workload.

The makespan outcome achieved for executing complex workload of varied size by HaSTE and MPAS-HMR is shown in Fig. 5. MPAS-HMR reduces makespan by 1.833%, 2.323%, and 6.226% when compared with HaSTE when workload size is 100MB, 200MB, and 400MB, respectively. From result obtained it can be state that MPAS-HMR improves makespan performance by 3.46% on an average when compared with HaSTE.

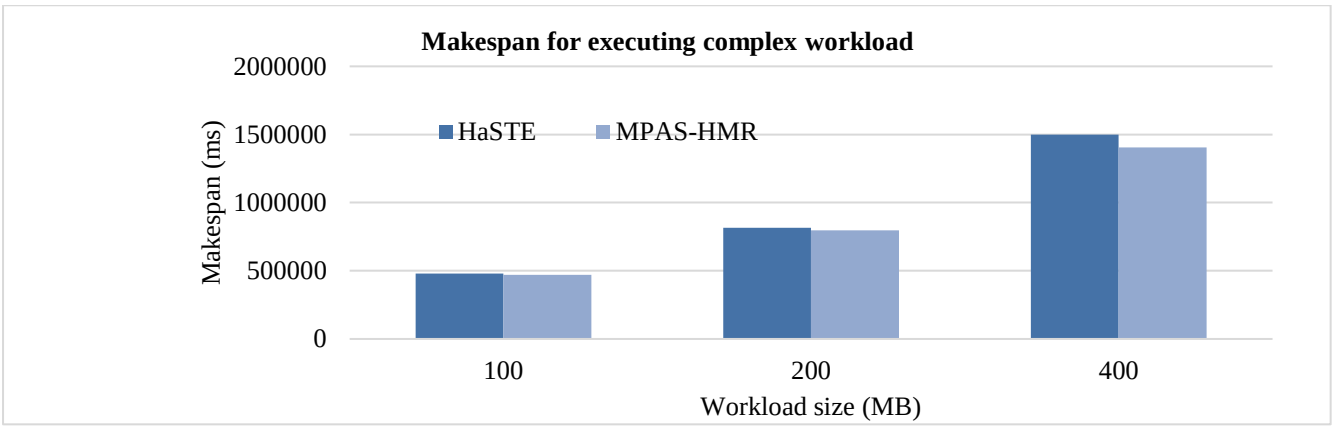


Figure 5. Makespan performance for executing complex workload.

The computational cost induced for executing complex workload of varied size by HaSTE and MPAS-HMR is shown in Fig. 6. MPAS-HMR reduces computational cost by 2.206%, 2.69%, and 6.58% when compared with HaSTE when workload size is 100MB, 200MB, and 400MB, respectively. From result obtained it can be state that MPAS-HMR reduce computation cost by 3.83% on an average when compared with HaSTE under varied workload scenarios.

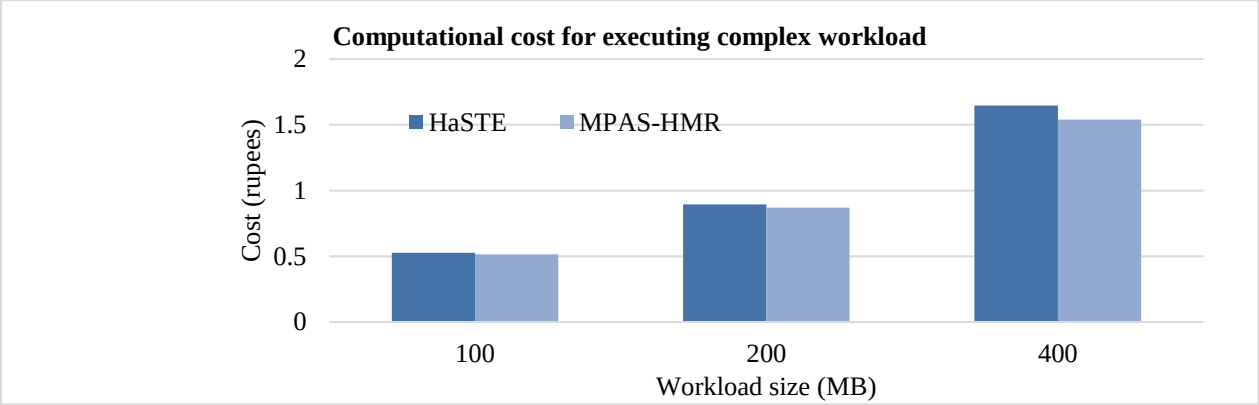


Fig. 6. Computational cost for executing complex workload.

IV. CONCLUSION

Managing memory resource is a challenging task. Since different phases of MapReduce job are executed concurrently. This paper presented memory and performance aware task scheduling adopting dynamic memory management technique and thread based task execution. Thus, uses memory resource and multi-core processing resource more efficiently when compared with existing HMR scheduler. Experiments are conducted using simple and complex workload. From result achieved it can be seen the MPAS-HMR reduce makespan and cost by 6.3% and 6.654% when compared with HaSTE for simple workload, respectively. Similarly, MPAS-HMR reduce makespan and cost by 3.46% and 3.83% when compared with HaSTE for complex workload, respectively. Thus, MPAS-HMR is efficient for running simple and complex iterative task. Though the MPAS-HMR achieves good result; still it is important to test the outcome considering heterogeneous workload. Further, need to evaluate how intermedia task failure affects makespan of scheduling model for HMR.

V. REFERENCES

[1]. Dean and Ghemawat (2008) MapReduce: Simplified Data Processing on Large Clusters, ACM Commun., vol. 51, no. 1, pp. 107–113.

[2]. Taura et al. (2003) Phoenix: a parallel programming model for accommodating dynamically joining/leaving resources, in SIGPLAN Not., , vol. 38, no. 10, pp. 216–229.

[3]. He et al. (2008) Mars: a MapReduce framework on graphics processors, in Proceedings of the 17th international conference on Parallel architectures and compilation techniques - PACT '08, p. 260.

[4]. Isard et al. (2007) Dryad: distributed data-parallel programs from sequential building blocks, ACM SIGOPS Oper. Syst. Rev., vol. 41, no. 3, pp. 59–72.

[5]. Kang et al. (2011) PEGASUS: Mining Peta-scale Graphs, Knowl. Inf. Syst., vol. 27, no. 2, pp. 303–325.

[6]. Lin et al. (2012) A Practical Performance Model for Hadoop MapReduce, in Cluster Computing Workshops (CLUSTER WORKSHOPS), 2012 IEEE International Conference on, pp. 231–239.

[7]. Cui et al. (2013) Modeling the Performance of MapReduce under Resource Contentions and Task Failures, in Cloud Computing Technology and Science (CloudCom), 2013 IEEE 5th International Conference on, vol. 1, pp. 158–163.

[8]. Khan et al. (2014) Data locality in Hadoop cluster systems, 2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Xiamen, pp. 720-724.

[9]. Xu et al. (2017) CRED: Cloud Right-Sizing with Execution Deadlines and Data Locality, in IEEE Transactions on Parallel and Distributed Systems, vol. 28, no. 12, pp. 3389-3400.

[10]. Alshammari et al. (2016) H2Hadoop: Improving Hadoop Performance using the Metadata of Related Jobs, in IEEE Transactions on Cloud Computing, vol. PP, no. 99, pp. 1-1.

[11]. Glushkova et al. (2017) MapReduce Performance Models for Hadoop 2.x, in Workshop Proceedings of the EDBT/ICDT 2017 Joint Conference, ISSN 1613-0073.

[12]. Ehsan et al. (2017) Cost-Efficient Tasks and Data Co-Scheduling with AffordHadoop, in IEEE Transactions on Cloud Computing, vol. PP, no. 99, pp.1-1.

[13]. Khan (2016), Hadoop Performance Modeling for Job Estimation and Resource Provisioning, in IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 2, pp. 441-454.

[14]. Apache (2014), Centralized Cache Management in HDFS.

[15]. Huang et al. (2016) YinMem: A distributed parallel indexed in-memory computation system for large scale data analytics, 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, pp. 214-222.

[16]. Zhang, J. et al.(2012), A Distributed Cache for Hadoop Distributed File System in Real-Time Cloud Services, in Proceedings of the 2012 ACM/IEEE 13th International Conference on Grid Computing. 2012, IEEE Computer Society. p. 12-21.

[17]. Longbin, L., et al.(2013) ShmStreaming: A Shared Memory Approach for Improving Hadoop Streaming Performance. in Advanced Information Networking and Applications (AINA), 2013 IEEE 27th International Conference on. 2013.

[18]. Kim et al. (2018) Selective I/O Bypass and Load Balancing Method for Write-Through SSD Caching in Big Data Analytics, in IEEE Transactions on Computers, vol. 67, no. 4, pp. 589-595.

[19]. Wu et al. (2017) Shadow: Exploiting the Power of Choice for Efficient Shuffling in MapReduce, IEEE 23rd International Conference on Parallel and Distributed Systems (ICPADS), Shenzhen, 2017, pp. 553-560, doi: 10.1109/ICPADS.2017.00078.

[20]. Yao et al. (2019) New Scheduling Algorithms for Improving Performance and Resource Utilization in Hadoop YARN Clusters," in IEEE Transactions on Cloud Computing, doi: 10.1109/TCC.2019.2894779.

[21]. Zheng et al. (2018) Joint Scheduling of Overlapping MapReduce Phases: Pair Jobs for Optimization. IEEE Transactions on Services Computing. PP. 1-1. 10.1109/TSC.2018.2875698.

[22]. Yang et al. (2019) Joint Optimization of MapReduce Scheduling and Network Policy in Hierarchical Data Centers, in IEEE Transactions on Cloud Computing, doi: 10.1109/TCC.2019.2961653.

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Analysis of Parallel Data Processing for The Sentimental Analysis Technique Using Spark and Machine Learning Algorithms

Kesavan M V[1], Dr.Josephine Prem Kumar[2], Manimegalai A[1], Shammi L[1]

[1]Assistant Professor, [2]Professor

[1]Department of Computer Science, East Point College of Engineering and Technology, Bengaluru, Karnataka, India

[2]Department of Computer Science, Cambridge Institute of Technology, Bengaluru, Karnataka, India

ABSTRACT

The market for big data is now expanding quickly. Finding a system that can store and manage a massive amount of data, then analysing that huge amount of data to mine the hidden knowledge, is a major task. This research proposed a comprehensive system for enhancing the performance of large data analysis. Both a big data storage environment and a quick big data processing engine using Apache Spark are included. About 11 Gigabytes of text data, gathered from various sources, are tested by the system for sentiment analysis. The Spark ML package already supports this system, which uses three separate machine learning (ML) techniques. The built-in model comprises of system programs developed in the Java and Scala computer languages.

Keywords: Apache Spark, Classification Algorithms, Amazon Datasets, Sentiment analysis, Hadoop HDFS, Big data, Machine Learning

I. INTRODUCTION

to The term of big data was presented and defined firstly in 2005 by Wigan and Clarke [1] as an immense volume of data that cannot be handled by old data management styles. Traditional business systems, internet/social networks, and internet of things, are the main sources for generating big data. At the end of the year 2020, the number of connected devices will be approximately one hundred billion. This significant increase resulted in a situation that the data is extremely big, and establishments are facing obstacles with handling and managing it. These recently evolving obstacles and challenges with the aspects of increasing data have been referred to as big data problems [2]. Hence, to overcome these problems it's crucial to understand big data analytics techniques and tools. These techniques and tools are preparing good manners and capabilities to manipulate all types of data, instead of using traditional systems that support only a small database [3]. To fulfill the institution's requirements, an efficient tool is important to treat and manage immense data size. In this regard, Apache Spark has appeared as a fast engine for processing large-scale data inside the computer memory. Through its in-memory processing environment, a Spark has been proven to be much faster than Apache Hadoop, specifically in iterative processing like machine learning (ML) programs. However, it can work across a diversity of nodes for parallel data processing. It has presented a new style for data

science where a wide range of data problems can be solved by a single processing engine [4]–[6]. Because Spark does not have its own data storage, well big data tool for handling a large amount of data is necessary. The most suitable tool for solving this problem is Apache Hadoop which consists of storagecalled Hadoop distributed file system(HDFS) [7].

The reason for the preference of Spark over the other big data tools is that the Spark ecosystem can offer a rich set of higher-level tools such as Spark SQL for SQL, MLlib for ML, Graph-X for graph processing and Spark Streaming for online processing. Also, Spark can support several programming languages which are Scala, Java, Python, and R [8]. ML package is the most interesting matter in Spark and consists of various types of ML algorithms such as classification, regression, clustering, collaborative filtering as well as model evaluation [9].

Supervised ML algorithms have been used in this study with Pipeline API in the Spark ML package. The Pipeline works to simplify the development and it is tuning multi-stage learning by providing a uniform set of high-level APIs [10]. In another term of meaning by using pipeline it can run a lot of algorithms according to a specific order. Therefore, it is useful for sequencing data pre-processing steps, which deliver cleaner and suitable data to the classifier algorithm. The dataset is passed through a group of preprocessing operations inside the ML pipeline and it is summarized in feature selection, data cleaning, data integration, tokenizing, stop word remover, stemming, and data representation [11]. Moreover, the whole procedure including Spark performance, Hadoop HDFS compatibility, and the effects of various utilized preprocessing steps have been tested through three types of algorithms which are logistic regression, support vector machine, and Naïve Bayes for binary opinion mining polarity (positive or negative) in the amazon product reviews.

II.APACHE HADOOP AND APACHE SPARK

Hadoop is an open-source structure written in Java programming language and designed for distributed storage and treating of very big datasets. It permits to store and analyze big data through multiple clients. Hadoop can scale up an individual's host to thousands of hosts and provide storage calculations for each one. Basically, the Hadoop framework isseparated into four parts HDFS, MapReduce, YARN, and Hadoop Common [12], [13].

HDFS

HDFS stands for Hadoop Distributed File System which divides the file systems into data and metadata. HDFS has two important benefits when compared with the traditional distributed file system. The first one is the great mistake tolerance and the second benefit it allows to use of big data sizes because the Hadoop clusters can remain data sets in petabytes [14]**.** The construction of HDFS is separated into one name-node and many data-nodes. A file in the HDFS is divided into a group of blocks and stored in the data-nodes. On the other hand the name-node is responsible for the block construction, termination, and replication [15].

MapReduce

MapReduce is a method for processing a large dataset stored in Hadoop HDFS. However, it permits the parallel processing of an enormous dataset. The MapReduce algorithm consists of two significant tasks, known as Map and Reduce[16].

YARN

Yet another resource negotiator (YARN) is the Hadoop cluster resource manager which means it handles the Hadoop cluster resources like memory and CPU. Fortunately, versions 2 and 3 of Hadoop with Yarn opens a new door for data treating environment [17].

Hadoop common

This part consists of Java libraries and some facilities that are required by other Hadoop parts. These libraries provide level abstractions for the OS, files system and necessary Java libraries and some scripts are compulsory to initialize Hadoop [15].

Apache Spark

Apache Spark is an open-source platform and in-memory (RAM) data processing. Spark allows fast processing of a huge data size leveraging distributed memory. It caches the data through multiple parallel operations, making it especially fast for parallel processing of distributed data with iterative algorithms. Spark can run multi-threaded lightweight jobs within the Java virtual machine (JVM) processes, supplying fast job start-up and parallel multi-core CPU utilization [18]. The tasks inside Spark accomplish multiple processes consecutively, in memory, only spilling to the local disk when required by memory limitations. It simplifies the management of multiple operations by offering a data pipeline technique. Spark characteristics are very suitable for big data ML and graph algorithms [19].

In addition, Spark was engineered from the bottom-up for performance, it can be multiple times faster than Hadoop for massive data processing by using in-memory calculating and many other optimizations. Besides the reason of time performance, many reasons make Apache spark more suitable for big data analysis over Apache Hadoop such as Hadoop only uses Map and Reduce operation, while Spark use Map, Reduce, Join, and Sample. However, Spark supports four programming environments which are Java, Scala, Python, and R [20], [21]. Using Scala of Spark increases the speed computation of the algorithms and completes them in less time as compared to Java furthermore, the favorites of Scala noticed in supervised ML algorithms such as regression and unsupervised ML algorithms like clustering [22].

Spark distributed environment and cluster managers

Hopefully, Spark is working in one node environment as well as in the multi-node environment. Without any doubt every multi-node environment needs a cluster manager, Spark supports four cluster manager types which are standalone cluster mode, Hadoop Yarn, Apache Mesos, and Kubernetes [18]. Spark architecture based on the distributedenvironment has three primary elements which are [23],[24]:

a. Driver program: it is the heart of the Job execution process in Spark and this element represents the slave node in a Spark cluster. The driver runs the application code that creates RDD's and then creates an object called Spark Contextthat administers and manages running applications

b. Cluster manager: this element is working for arranging the application workflow that allocated by the driver program to the workers. However, it operates all the resources in the cluster and brings back their situation to the driver program. In another word, the cluster manager is controlling the whole communication between the master node and the slaveswhen they run an application

c. Worker nodes: every worker node denotes a container of one operation throughout the Spark program execution. In another term of meaning, each worker-node has its executors and every executor will run several jobs.

In the Spark distributed environment, the driver program runs in its own Java process. These drivers communicate with a potentially huge number of distributed workers who are called also executors. Every executor is a single java process. A Spark application is a mixture of the driver and its executors. Spark application is running on a group of machines with the assistance of the utilized cluster manager [18].

Spark data access and data structure

One of the greatest advantages of Apache Spark is accessing/reading the data from multiple places such as HDFS, Mesos, Mongo DB, Cassandra, H-Base, and Amazon S3. Similarly, Spark can store/write the data in all the mentioned data storages which means it has a diversity of data reading and writing from various data sources [25]. Also, the data structure of Apache Spark fundamentally consists of three types of data which are [18], [26], [27]:

d. Resilient distributed datasets (RDD): spark uses a particular data structure known as RDD which is a logical

collection of data and separated over machines. RDD is Spark's primary abstraction, which is a fault- tolerant collection of elements that can be worked in parallel

e. Data frame (DF): it is a dataset organized into named columns or a collection of distributed records. DF is exactly such as RDD but, it is shaped into named columns with covering the characteristics of Spark SQL's execution. It is conceptually like a table in a relational database with better optimizations

- Dataset: it is a distributed collection of data. Dataset is a new interface inserted in Spark 1.6 that offers the benefits ofRDDs with the benefits of Spark SQL's optimized

III. RELATED WORK

A tremendous number of researches are printed on the subject of big data processing recently by utilizing Hadoop and Spark. There have been several approaches to analyzing big data. In this section, the focus only will be on the credible movements and contributions of this field. M. Assefi et al, 2017 [28] explored some views for growing the form of the Apache Spark MLlib 2.0 as an open-source, accessible and achieve ML tests that related to the real world to inspect the attribute characteristics. Also presented a comparison among spark and Weka with proving the advantages of spark over the Weka in many sides like the performance and it is efficient dealing with a huge amount of data. on the other hand, Weka is good for simple users with its GUI and the diversity of algorithms that already exist in it.

Yan *et al*. [29] discovered a micro blog sentiment classification scheme with paralleled support vector machine in the spark multi-node environment. They rose the accuracy by feature space evolution and tuning the parameters. Moreover, the execution speed is risen

with Apache Spark as well as to the speed of support vector machines (SVM) with radial basis function (RBF). However, the capability of Spark is completely utilized since the dataset is extremely large. Al-Saqqa *et al*.[30] discussed Spark's MLlib for hiring it in the classification of sentiment big data scale. They found that SVM is better than the other classifiers in the matter of performance.

Barznji *et al*. [31] talked about sentiment analysis utilizing the algorithms of ML such as Naïve Bayes and SVM for analyzing the text with the benefits of the huge capabilities of Apache Spark. They found that the SVM is more accurate in the condition of total average. Finally, Symeonidis *et al.* [32] tested some important pre-processing techniques and evaluated them in two datasets. Each method tested the accuracy in four ML algorithms. Furthermore, the study was implemented on all, as well as on the high-performance methods, in order to find method interactions. Their investigations present that a few methods supply better outcomes in both datasets utilized for Twitter sentiment analysis, while other methods reduce the accuracy suchas substituting slang and spelling alteration.

Dataset

Sentiment or opinion analysis of product-reviews data denotes the feeling and attitude of the individual on a product and it is available in online sources such as the Amazon product reviews site. In this work, three different datasets about Amazon Reviews have been used because it is the most wildly utilized for opinion analysis in an interactive way. Amazon Reviews is a platform, where customers can post comments on any product and ask/answer questions or share their own opinions. This work will not be limited to concentrating only on the topic of a particular review.

No.	Name of the dataset	Size	No. of fields (attributes)	No. of rows
1.	Amazon reviews: kindle store category	685 MB	Nine	982,899
2.	Amazon reviews for sentiment analysis	1.6 GB	Two	3,607,482

3.	Web data: Amazon movie reviews	8.69 GB	Eight	7,903,890
4.	Total size	10.9 GB	Two	12,494,271

Table 1.Datasets with their original size

Three data reviews about different categories have been collected from the Kaggle Repository.

The first dataset is called Amazon reviews-Kindle store category, the second dataset called Amazon reviews for sentiment analysis and the third dataset called web data: Amazon movie reviews. Table 1 shows additional details on the datasets.

IV.THE PROPOSED SYSTEM ARCHITECTURE

The proposed system of this work is constructed on a VMware Workstation software version 15.0.2, which is used to handle Linux- Debian 9, 64-bits as a guest operating system (O.S). The host operating system of the VMware program is Windows 10, 64-bits. Besides, various big data tools have been installed on the guest OS such as Hadoop and Spark. Distributed environments for both Spark and Hadoop are used to gain parallel data processing utilizing three nodes: one master and two slaves (with the capability of expansion for future works). Then Spark needs to be combined with Hadoop to read and write data from and to HDFS. This combination will enrich the data processing capabilities. In other words, Hadoop works to distribute the data across multiple nodes to be seen by all the Spark nodes which work to process the data using in parallel (parallel data processing). In general, the proposed system architecture is explained in Figure 1.

The workflow of the proposed system is consisting of many steps starting from data collection and ending with model testing, all these steps are illustrated by the following four points:

a. Feature selection: a program was written in a java programming language to read the dataset and select the required columns (features) which are text and label from many unimportant columns.

b. Data cleaning-1: a program was written in a java programming language to clean the data by removing unwanted characters and commas within the text column and keeping the sole comma that separates between text and label for each comment.

c. Data integration: this step is divided into three types; each type will be processed separately in the upcoming steps. Type (A), integrated dataset: the three datasets are combined and saved in a single dataset file.

   Type (B), datasets with equal number of comments: A single dataset file (Type A) was divided into three new files with the same number of comments for each one of them.

   Type (C), normal size of datasets: loading and processing the original dataset files with their original size without any integration or division.

d. Data pre-processing and prediction: in this step, the main program for reading and processing data was written in the Scala programming language. Two approaches are used here: The first approach (central data processing) using for reading the integrated dataset (single dataset file-Type A) from the local disk with one node. While, the second one (distributed data processing) using for reading the data from HDFS within three nodes. The distributed approach read all the three types of data (Type A, Type B, and Type C) separately to evaluate the accuracy and performance of each one of them. Then applying data pre-processing stages: datacleaning-2, tokenizing, stop-words remover, and ineffective-words remover which works to remove many words that do not have any effects on the polarity type (positive or negative). After that using the stemming and finally feature extraction (hashing TF and IDF) for changing the texts into vectors.

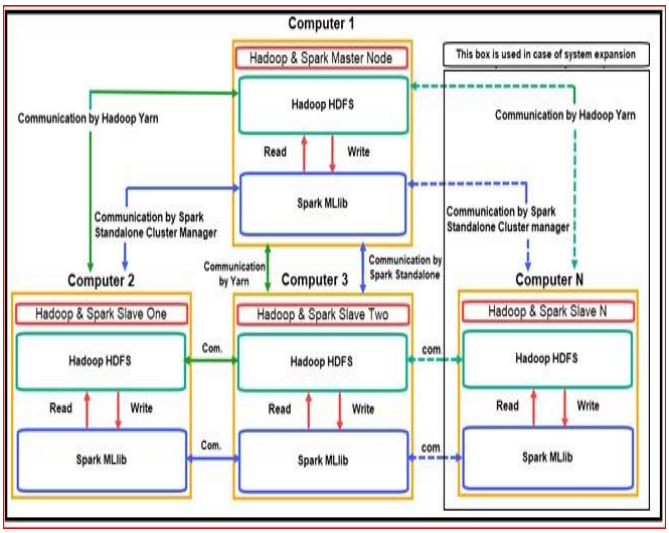


Figure 1. The proposed system architecture

Table 2 shows the characteristics of the datasets after the first three steps of data preprocessing which are feature selection, , data cleaning-1, and data integration. The whole procedures for the above four practical points are shown in

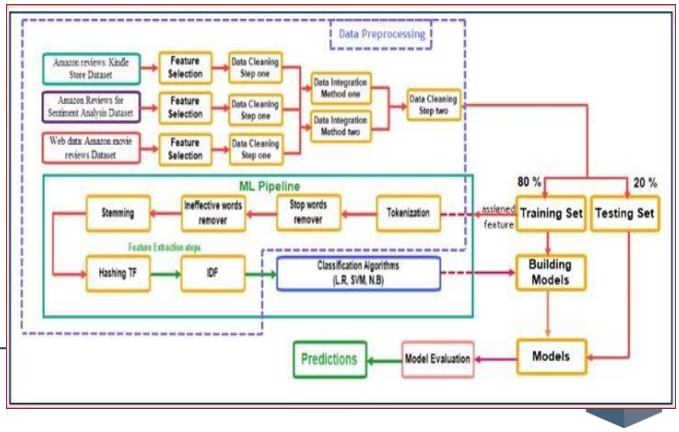
Table 2. Characteristics of the datasets after the first three steps of data preprocessing

	Name of the dataset	Size	No. of fields	No. of rows
1.	n reviews: Kindle ategory	570 MB	Two	982,899
2.	Amazon reviews for sentiment analysis	1.37 GB	Two	3,607,482
3.	Web data: Amazon movie reviews	6.4 GB	Two	7,903,890
4.	Aggregation files (integration of the three datasets)	8.35	Two	12,494,271

Implementation

The Implementation of this work has been done in a very efficient manner which is convenient to use this system for analyzing any data size. The final utilized dataset size was reduced from 10.9 GB to 8.3 GB after the first two pre-processing steps as well the integration process. The implementation is divided into two approaches, approach one is central data processing (single node data processing) and approach two is distributed data processing (multi nodes data processing).

Figure 2. Proposed system implementation steps



moreover, utilizing three types of algorithms for classifying the texts which are logistic regression, SVM, and Naïve Bayes. All the implementation steps starting from pre-processing and ending with the classifiers must be executed according to a particular order. So, the pipeline is required for ordering procedures. After the pipeline stage, the data is divided into 80% for the Training set and 20% for the testing set. Finally, testing the constructed model based on the remaining 20% of data and then getting positive and negative results.

Figure 3 shows the proposed system implementation steps. Furthermore, in approach two, multi nodes data processing, the same implementation steps are applied with utilizing Hadoop HDFS for distributing data into blocks across the three nodes. As well As using Spark distributed environment for parallel data processing across the three slaves. Figure 4 shows the screenshot of Spark distributed web console monitoring.

IV. RESULTS AND DISCUSSION

As mentioned before, three classification algorithms were utilized in this work which are logistic regression, SVM, and Naïve Bayes. For each algorithm, the measures of accuracy, precision, recall, f-measure as well as execution time have been computed for both mentioned approaches.

Central data processing (approach one)

The results of all the mentioned measures as well as the comparison between the algorithms have been illustrated in Table 3. The outcomes of binary sentiment analysis using logistic regression and SVM classifiers obtained an excellent rate of accuracy on training data and Naïve Bayes accuracy result is very good. From these results, it can conclude that the utilized pre-processing steps produced a significant positive impact on classification accuracy and execution time.

Table 3. Central data processing results

Classi. algo.	Accuracy	Precision	Recall	F-meas.	Exec. time in min.
LR	90.7%	90.5 %	90. 7%	90.4%	146.7
SVM	90.0%	89.8 %	90. 0%	89.6%	684.1
NB	80.8%	84.9 %	80. 8%	81.9%	145.8

Distributed data processing (approach two)

A few manipulations process on the datasets has been done. This manipulations process has resulted three types of data as mentioned before which are: type (A), (B), and (C). This manipulations process is to gain better accuracy and performance among all the utilized types. However, to check if the system is more interacting with one huge dataset file or with multiple datasets that have the same size as the huge file. The outcomes present that all the utilized data types have approximately the same measures result as shown in Figures 3 and 4.

After performing both approaches central and distributed data processing, now showing some brief comparisons between them are necessary to check the best approach.

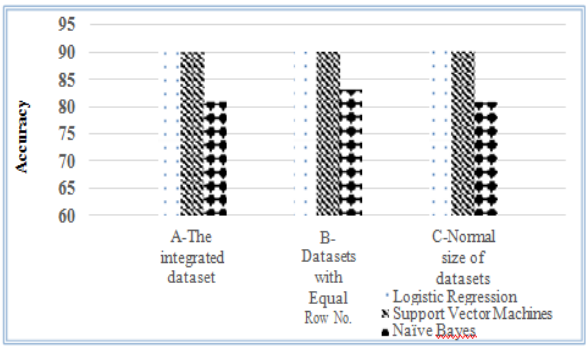


Figure 3. Accuracy comparison between utilized data type in approach two

The comparisons focus on-time performance as well as on the previously obtained measures such as accuracy and F-measure. Because the acquired time performance and the other measures for all the utilized data types (A, B, and C) in the distributed data processing are the same, the comparison will be done between Central data processing and Distributed data processing. The experimental results show that the learning times for all the classifiers are reduced approximately into half by using the distributed system approach as compared with the central approach.

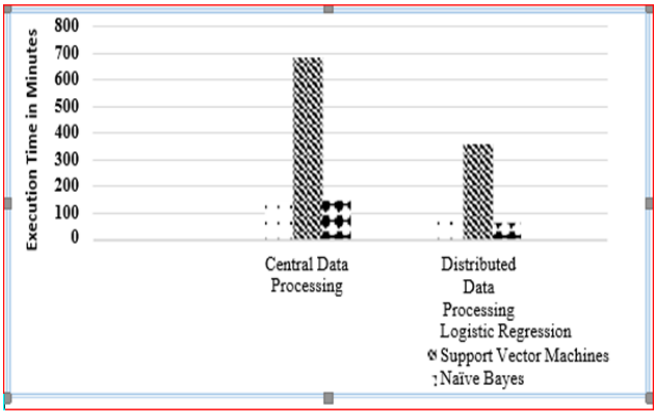


Figure 4. Execution time differences between the approaches

Another measure obtained from all the experiment tests was the amount of time taken for constructing the models in central and all the distributed types. Naïve Bayes required the smallest amount of time to complete the model building and then logistic regression. Whereas, the learning time for the SVM algorithm was very long even in a distributed manner. It is obvious that in both central and distributed manners the highest accuracy and F-measure rates achieved by logistic regression algorithm and then SVM algorithm.

Whereas the Naïve Bayes algorithm scored the lowest accuracy and F- measure rates, in other words, it has the worst results as compared to the other classifiers.

V. CONCLUSION

Undoubtedly big data is one of the major drivers of digitization. In the technical revolution, every single word/number might be useful information and has its benefits for the foundation's progress. The main goal of this work is to build a big data prototype system that consists of two crucial big data requirements which are distributed data storage for handling any size of big data as well as applying parallel data processing across that distributed storage for decreasing the execution time as much as possible. For gaining these two goals Apache Hadoop and Spark have been used in the system construction. The system has been tested with 11 Gigabytes of big text data for sentiment analysis. The datasets are collected from three different Amazon customer reviews and then aggregated in one huge file. Unlike small data, big data needs various data pre- processing steps to obtain good time performance and accuracy results. The proposed preprocessing methods of this work reduced the overall dataset size, only in the first two steps of the utilized pre- processing the size of the aggregated dataset file reduced to 8.3 Gigabytes which consists of about 12,494,271 review comments with binary labels (positive and negative). This reduction leads to performing and applying the algorithms in less time with gaining a better accuracy rate.

Two approaches are utilized in this system, central data processing(one-node) and distributed data processing (multi-node). Distributed processing approach outperformed the central processing approach and it is approximately reduced the execution-time of the model construction into half as compared with the central approach in all the utilized algorithms with acquiring the same accuracy ratio. The system

programs have been written in Java and Scala programming languages and the constructed model consists of the classification algorithms as well as the preprocessing steps in a figure of ML Pipeline stages. Three types of algorithms have been used and the experiments indicated that the Logistic Regression out performed both SVM and Naïve Bayes classifiers in both utilized approaches. The final results showed that the system can treat a massive size of data with fast execution time especially in the distributed data processing approach. However, it can gain a very good accuracy depending on the various applied phases of the data pre-processing. In addition, the system tests were presented that the used manners of dataset manipulations such as separated datasets processing (multi-files) or integrated dataset processing (one-file) do not influence on the accuracy and time performance results.

II. REFERENCES

[1]. M. R. Wigan and R. Clarke, "Big Data's Big Unintended Consequences," IEEE Computer Society, vol. 46, no. 6, p. 46–53, 2013, doi: 10.1109/MC.2013.195.

[2]. P. Géczy, "Big Data Management: Relational Framework," Review of Business & Finance Studies, vol. 6, no. 3, pp. 21–30, 2015.

[3]. S. Alkatheri, S. Anwar Abbas, and M. A. Siddiqui, "A Comparative Study of Big Data Frameworks," International Journal of Computer Science and Information Security (IJCSIS), vol. 17, no. 1, pp. 66–73, January 2019.

[4]. S. Gopalani and R. Arora, "Comparing Apache Spark and Map Reduce with Performance Analysis using K-Means," International Journal of Computer Applications (0975 – 8887), vol. 113, no. 1, pp. 8–11, March 2015, doi: 10.1.1.695.1639.

[5]. L. R. Nair, S. D. Shetty, and S. D. Shetty, "Applying spark based machine learning model on streaming big data for health status prediction," Computers and Electrical Engineering, vol. 65, pp. 393–399, 2017, doi: 110.1016/j.compeleceng.2017.03.009.

[6]. S. Ra, B. Ganesh H.B., S. Kumar S, P. Poornachandran, Soman K.P., "Apache Spark a Big Data Analytics Platform for Smart Grid,"Procedia Technology, vol. 21, pp. 171–178, 2015, doi: 10.1016/j.protcy.2015.10.085.

[7]. D. Graux, L. Jachiet, P. Genev`es, and N. Laya¨ıda, "SPARQLGX: Efficient Distributed Evaluation of SPARQL with Apache Spark," in The 15th International Semantic Web Conference - the Semantic Web – ISWC 2016. ISWC 2016, pp. 17–21 October 2016, doi: 10.1007/978-3-319-46547-0_9.

[8]. M. A. Khan, Md. R. Karim, and Y. Kim, "A Two-Stage Big Data Analytics Framework with Real World Applications Using Spark Machine Learning and Long Short-Term Memory Network," Symmetry, vol. 10, no. 10, pp. 1–19, 2018, doi: 10.3390/sym10100485.

[9]. S. Harifi, E. Byagowi, and M. Khalilian, "Comparative Study of Apache Spark MLlib Clustering Algorithms," in International Conference on Data Mining and Big Data-Second International Conference, DMBD, pp. 61–73, 2017, doi: 10.1007/978-3-319- 61845-6_7.

[10]. X. Meng et al, "MLlib: Machine Learning in Apache Spark," Journal of Machine Learning Research, vol. 17, no. 1, pp. 1235– 1241, 2016, doi: 10.5555/2946645.2946679.

[11]. O. Faker and E. Dogdu, "Intrusion Detection Using Big Data and Deep Learning Techniques," in ACM SE '19 Proceedings of the 2019 ACM Southeast Conference, pp. 86–93, 2019, doi: 10.1145/3299815.3314439.

[12]. "Apache Hadoop," Apache software foundation, [Online]. Available: https://hadoop.apache.org/. [Accessed 6-3-2019].

[13]. Y. Mjhool, A. H. Alhilali, and S. Al-Augby, "A proposed architecture of big educational data using Hadoop at the University of Kufa," International Journal of Electrical and Computer Engineering (IJECE), vol. 9, no. 6, pp. 4970– 4978, 2019, doi: 10.11591/ijece.v9i6.pp4970-4978.

[14]. T. Ruzgas, K. Jakubėlienė, and A. Buivytė, "Big Data Mining and Knowledge Discovery," Journal of Communications Technology, Electronics and Computer Science, no. 9, pp. 5–9, 2016, doi: 10.22385/jctecs.v9i0.134.

[15]. C. Kaushal and D. Koundal, "Recent trends in big data using Hadoop," International Journal of Informatics and Communication Technology (IJ-ICT), vol. 8, no. 1, pp. 39–49, April 2019, doi: 10.11591/ijict.v8i1.pp39-49.

[16]. A. Gupta, G. Aggarwal, and A. Kumar, "Hadoop, Map Reduce and HDFS-A Review," International Journal of Scientific & Engineering Research, vol. 8, no. 12, pp. 82–84, December 2017.

[17]. Y. Perwej, B. Kerim, M. Sirelkhtem, and O. E. Sheta, "An Empirical Exploration of the Yarn in Big Data,International Journal of Applied Information Systems (IJAIS), vol. 12, no. 9, pp. 19–29, December 2017, doi: 10.5120/ijais2017451730.

[18]. "Apache Spark," Apache software foundation, 24 Feb 2019. [Online]. Available: http://spark.apache.org/.

[19]. K. Wang and M. M. Hasan Khan, "Performance Prediction for Apache Spark Platform," in 2015 IEEE 17th International Conference on High Performance Computing and Communications (HPCC), pp. 166–173, 2015, doi: 10.1109/HPCC-CSS- ICESS.2015.246

[20]. J. P. Verma and A. Patel, "Comparison of MapReduce and Spark Programming Frameworks for Big Data Analytics on HDFS,"IJCSC, vol. 7, no. 2, pp. 80–84, 2016.Verma, A. H. Mansuri, and N. Jain, "Big data management processing with Hadoop MapReduce and spark technology: A comparison," Symposium on Colossal Data Analysis and Networking (CDAN), pp. 1–4, 2016, doi: 10.1109/CDAN.2016.7570891

[21]. H. K. Omar and A. K. Jumaa, "Big Data Analysis Using Apache Spark MLlib and Hadoop HDFS with Scala and Java," Kurdistan Journal of Applied Research (KJAR), vol. 4, no. 1, pp. 7–14, June 2019, doi: 10.24017/science.2019.1.2.

[22]. W. Inoubli, S. Aridhi, H. Mezni, M. Maddouri, and E. M. Nguifo, "An experimental survey on big data frameworks," Future Generation Computer Systems, vol. 86, pp. 546–564, 2018, doi: 10.1016/j.future.2018.04.032.

[23]. R. K. Chunduri and A. K. Cherukuri, "Scalable formal concept analysis algorithm for large datasets using Spark," Journal of Ambient Intelligence and Humanized Computing-Springer Berlin Heidelberg, vol. 10, pp. 4283–4303, 2018, doi: 10.1007/s12652-018-1105-8.

[24]. U. R. Pol, "Big Data Analysis: Comparision of Hadoop MapReduce and Apache Spark," International Journal of Engineering Science and Computing (IJESC), vol. 6, no. 6, pp. 6389–6391, 2016, doi: 10.4010/2016.1535.

[25]. J. Peloton, C. Arnault, and S. Plaszczynski, "FITS Data Source for Apache Spark," Computing and Software for Big Science- Springer, vol. 2, no. 7, pp. 1–9, October 2018, doi: 10.1007/s41781-018-0014-z.

[26]. Sassi, S. Anter, and A. Bekkhoucha, "A spark-based paraldistributed posterior decoding algorithm for big data hidden Markov models decoding problem," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 10, no. 3, pp. 789–800, 2021, doi: 10.11591/ijai.v10.i3.pp789-800.

[27]. M. Assefi, E. Behravesh, G. Liu, and A. P. Tafti, "Big data machine learning using apache spark MLIib," in 2017 IEEE International Conference on Big Data , pp. 3492–3498, 2017, doi: 10.1109/BigData.2017.8258338.

[28]. B. Yan, Z. Yang, Y. Ren, X. Tan, and E. Liu, "Microblog Sentiment Classification using Parallel SVM in Apache Spark," in 2017 IEEE 6th International Congress on Big Data, pp. 282–288, 2017, doi: 10.1109/BigDataCongress.2017.43.

[29]. S. Al-Saqqaa, G. Al-Naymata, and A. Awajan, "A Large-Scale Sentiment Data Classification for Online Reviews Under Apache Spark," Procedia Computer Science, vol. 141, pp. 183–189, 2018, doi: 10.1016/j.procs.2018.10.166.

[30]. Al-Barznji and A. Atanassov, "BIG DATA SENTIMENT ANALYSIS USING MACHINE LEARNING ALGORITHMS," in Proceedings of 26th International Symposium "Control of Energy, Industrial and Ecological Systems, pp. 53–58, May 2018.

[31]. S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactionsfor Twitter sentiment analysis," Expert Systems with Applications, vol. 110, pp. 298–310, 2018, doi: 10.1016/j.eswa.2018.06.022.

[32]. B. Srigiriraju, "Amazon reviews: Kindle Store Category," [Online]. Available at: https://www.kaggle.com/bharadwaj6/kindle- reviews. [Accessed 22-7-2019].

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Traffic Light Cycle Control Using Deep Reinforcement Learning

Madhu shree R*[1], Dr. Heena Kosur[1], Ammu bhuvana D[1], Maharaj S B[2]
*[1]Department of Computer Science and Engineering, East Point College of Engineering and Technology, Bangalore, Karnataka, India
[2]Department of Mechanical Engineer, Bangalore, Karnataka, India

ABSTRACT
Numerous issues are brought on by a traffic light cycle control, including energy waste and significant delays for moving traffic. The issue cannot be solved using the current traffic signal control technologies and approaches. Vehicle traffic is becoming more and more congested every day. Since there are more cars on a certain side of the road, it is necessary to redesign the traffic light system. Static Signal switching is capable of handling and monitoring signals in real-time. We provide a paradigm of deep reinforcement learning to regulate the cycle of traffic lights, traffic data and gridding the entire intersection. For the model in a real-world scenario, this study addresses the control of the vehicle and attempts to find a solution to the current issue. The YOLO algorithm has been pre-trained, and the fundamental simulation, vehicle identification, tracking, and vehicle counting work is preserved for future signal processing requests. In the initial stages, CCTV cameras are used to record the images. The "Traffic Control Interface" (TraCI) is a Simulation of Urban Mobility (SUMO) interface gives users access to a live retrieving values for simulated objects in road traffic simulation, and handles single and multiple-intersection cases.

Keywords: Deep Learning, Video Processing, Image Processing, Object Detection, Convolutional Neural Network.

I. INTRODUCTION

All cities now exhibit instability and display complicated dynamics as a result of the quick change that is occurring in this direction [1]. Surviving in such a quicker landscape has both advantages and disadvantages. However, safety is compromised as a result of this speed. The number of automobiles on the road is growing significantly daily, and hence, accidents are more common. This outlines the concept of autonomous or self-driving vehicles. A self-driving automobile must be capable to comprehend and follow the adopt them correctly in order to function properly. [2] In order to maximize the efficiency and convenience of Intelligent Mobility Systems, which is one of the key components of Smart Cities, VANet techniques using vehicular communication taken into consideration (ITS). A specialized and distinct kind of vehicular networks are vehicular ad hoc networks. It is based on IEEE 802.11p, which is the only lexicon for describing data transmission [3]. Consequently, the need for effective traffic sign detection tools has increased. Road accidents can be brought on by careless driving and failing to pay attention to the surroundings. Furthermore, certain

weather conditions could impair drivers' ability to see particular signs and result in catastrophic accidents. Traffic lights are typically used to regulate intersections along busy or major highways. However, their ineffective regulation results in a number of issues, including significant energy waste and traveler delays. Even worse, it might result in car accidents.. The detection of vehicles play a major role for the segmentation of an image in the first steps of video processing. The accuracy and efficiency in each vehicles is needed for tracing, gesture and detection for the eventually processing of vehicles. The vehicles detection counting plays a major role in system that need to manage and control traffic within the metropolis. The major aim is detection, counting, tracking and speed estimation of each car with maximum accuracy on roads and highways. The concept applied here is to apply the front objects, Convolutional Neural Network takes the car as input to the video for processing the accurate result for counting vehicles. The purpose of using convolutional neural networks has made a great achievement in the vehicle detection. Convolutional Neural Network has a potential to acquire skill of pictures feature that has an ability to perform multitasks, which are regression and classification. YOLO is abbreviated as “You Only Look Once” This programme uses a convolutional neural network (CNN) to track and identify different items in a picture. It can identify objects like cars. This approach produces precise results with few background errors and only needs one forward propagation across a neural network The frames can be divided using video or live footage. CNN algorithm is used to track and detects the car when it moves into the Region of Interest(ROI). We can easily recognize and categorize the items in photos, however this is a challenging operation for computers to perform in real time. Though the trained algorithm is powerful enough, the object can still be recognized and accurately identified even if it is only partially visible. Many issues have evolved in modern life that we must deal with, such as traffic accidents and congestion. One of the best contemporary approaches being used by countries to improve the traffic system is the use of this data, which may be critical in many surveys and is essential to the traffic management system. This method allows us to the most cars should be counted precise manner possible. Using this way The most accurate methods of counting the number of automobiles are available. The main The purpose of this endeavor is to create a fast and reliable solution to the traffic congestion problem. The solution must be intelligent and has decision-making capabilities depending on the different situations in the different lanes. There are different phases of the proposed model. The first phase is video processing. The continuous video stream from the traffic camera situated at the traffic signal junction is fed to the model. During the video processing, the frames are extracted from the video input. There are hundreds of frames present in a stream of few seconds, so it is not possible to process each frame and it is not required because there is no such significant change in the vehicle density between two consecutive frames. Therefore, each frame after a particular time interval is taken. After this phase, the processed frame is fed to the model for object detection. There are two separate models developed for object detection using each framework i.e., YOLOv3 and YOLOv5.

Figure 1: Traffic road cycle control

The neural network extracts the features and identifies the object using bounding boxes. After object detection, The model predicts the class of the vehicle from a car, bike, truck, etc. The total vehicle count of a lane is passed to the traffic signal timer algorithm. The algorithm considers the density of all the other lanes and calculates the relativity between them. Depending on which, it classifies the lane in either low, medium, or high vehicle density class, it decides the green signal timer for a particular lane. The scope of this project is the vehicle detection on roads uses the deep learning technique. Earlier the datasets used for this project was small which falls less in multiple aspects. Firstly, time duration for the road using the vehicle datasets are small in terms. Further, camera is used by the researchers which is useful and supportive for the vehicle detection by using deep learning method which has the limited scope as contrast to the input composed by the transportation authorities.

II. LITERATURE REVIEW

Priyanka Bhamare et.al, [1] tells about the method that is used for vehicle recognition technology used in traffic signals. To detect the cars Smart cameras are integrated into the system to make it work. The use of the automobile detection in the video by CNN algorithm and the images are trained and tested using this algorithm. Indrabayu et.al,[2]talks about the tracking and vehicle detection, here Kalman flite along with Gaussian mixture is used to perform the task. For detection of vehicles the video input is taken from the signals. Li xun et.al,[3]says that the vehicle condition methods that are available in order to extract are not sufficient, an efficient procedure method should be used in sequence to process the model. Ahmad Aliet.al, [4]In this paper low-cost camera is used for the basic idea of building an environment and the algorithm was used for concerning and control the vehicles that are passing on the lanes. Krishnamoorthy et.al, [5]describes about the traffic that are carried out in many countries. In this they used the CCTV cameras to automate the traffic signals connected over the internet to survey various junction at the roads. Jess Tryon et.al, [6]says about the traffic in Philippines that affect the residents and industries sectors. Here the intelligent transportation system is used based on the real-time traffic in cities. The traffic can be controlled by deploying CCTV at every lanes and the data is sent to Raspberry for calculating image density. Asra Aslam et.al [7] finds an enormous amount of sensing data devices that collects the information over the internet of things. The services that are provided by IOT are sensing, networking, service and application-level services. Poonam A et.al [8] says about the object detection is most important and challenging branches that are used in computer vision. the rapid development of deep learning methods is used for detection tasks after that variety of object methods are used for detecting covering with the onestage and two-stage detectors. Imran Mahmud et.al, [9] focused on the traffic density of the capital of Bangladesh, dhaka. Here its very tough for the traffic police to controlling the traffic who break the laws and jump the traffic signals. Lv Ning et.al, [10]tells about the vehicle counting from an aerial video(UAV) in traffic monitoring, which can be deployed in areas by collecting the data from camera as a visual sensors. Akoum et.al, [11] had implemented a Smart traffic controllers used real-time images as well as a filtering technique to remove only the waste objects from the images. shows the objects like cars. Nadia Baha, [12] says about the sensor that are used to detect the obstacles utilising real-time stereo vision in both indoor and outdoor environments. The method combines with thresholding and accumulating techniques to cluster and detect using a disparity map to identify obstacles.

III. PROBLEM STATEMENT

In today's world the theft of vehicles are increasing more and reported every year. If the vehicles that are being stolen are not recovered early, then it has been sold or burned. It is very tough to find the stolen vehicle using this paper we can detect and track, which protects the vehicles. Traffic surveillance is commonly used to detect and track moving vehicles. The tracking of vehicle is very versatile and can be refined to do additional tasks such as vehicle tracking and counting. Figure2The cost estimation using this technology will be very less, Video and image filtering has been used for traffic observation, analysis and monitoring the traffic condition in many cities and urban areas. All process attempt to increase the accuracy and decrease cost for hardware implementation. The main objective of this paper is detection of vehicles, tracking and estimating the speed accuracy on highways and in small lanes using convolutional neural network algorithm. YOLO algorithm detects the static vehicles and ignore the shadows and reflection which results in traffic on highways.

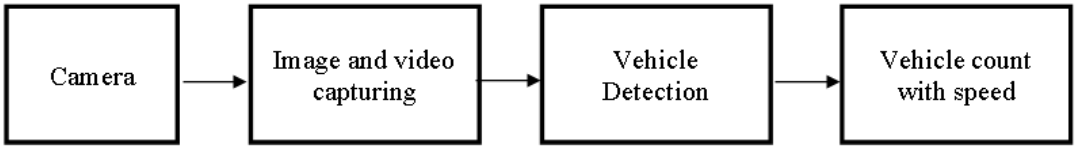


Figure 2: Overview of the proposed system

IV. METHODOLOGY

The system aims to provide a solution for managing the vehicle in real time scenario. To perform a specific task the pre-trained model YOLO is used to detect the objects in the sense of vehicles, Image processing is employed for speed calculation and count of vehicles. At an initial level the model is ready with all types of camera even the low cost including the surveillance camera. The model will receive a captured image for the vehicle detection purpose, after that using single camera the whole process will be repeated for all four sides of the road. To capture the image the camera will be fixed at one place. The vehicles will be detected and counted which are captured inside the given Region of Interest(ROI). The image size remains constant after being captured. The object tracking and detection plays an important role using OpenCV library within the rectangular box inside the vehicles. In all the sides of road the count can be obtained after passing the input to the data storage. The different images obtained from the count of vehicles can predict the result by using the computers and laptops. The system have a threshold unit value that are fixed and after that if the result from the images limit within the threshold then simple switching is started for every signal to control the vehicles in the traffic, after that if there is more traffic in one direction then the vehicle will be sent to another direction by switching the traffic signals. The methodology of proposed system is illustrated in the figure 3.

1) **Video accession** : This method is used for acquiring the video through any one the devices video camera, smart-phone camera, universal serial bus camera etc.
2) **Frame transformation:** Frame transformation is a process of converting the frames one by one using the given video. Once the videotape that capture the videos, those videos are converted into fixture and suitable type of process can be done accordingly to the frame.

3) **Pre-Processing** : The Pre-Processing method is added for the video which is used to decrease the sound. some of the techniques of preprocessing are smooth, dilate, erode and median etc.
4) **Background Modelling**: The pre-processing is used to create an ideal background according to the environmental changes. The image subtraction operations can be performed by using the background modelling. There are two types of techniques recursive or non-recursive that are used in background modelling.
5) **Background Subtraction:** he background subtraction is the main move for processing. First in this any changes in the image region from the background of the model are predicted, then the pixels are arranged in the given regions after that the changes are made for further processing. The labelling algorithm is used to connect the regions of the component.
6) **Post-Processing**: To improve the results of prepprocessing which is used after subtraction and background modelling, the foreground mask is implemented in pre-processing techniques.
7) **Foreground Extraction** : This is the last step for processing the computer vision and processing the image, whose aim is to detect changes in sequences of images using subtraction of the background method for further processing.

Video Processing: In this first part of the traffic control system, the inputs will be accepted in separate four individual videos of each lane in focus. These videos might come in any resolution or colour format. The first task in this segment of the solution will be to update the resolution of the input videos to make all four input uniform and consistent for the detection model. The videos are adjusted to a specific resolution of 416 by 416 pixels each and the colour formats of the video will be adjusted in the RGB (Red, Green, Blue) colour format. Any videos which might get sent in other colour schemas like CMYK (Cyan, Magenta, Yellow, black) or HSV (Hue, Saturation, Value) will be converted to RGB in a 3- dimensional array structure, containing 3, 2-dimensional matrices of each colour component value in the video frames. Finally, these videos will be chopped down to some selected frames based on a certain interval.

Object Detection: The next phase of the proposed solution is to apply object detection to the received frames from the previous stage. Here the frames received will be passed to the object detection model of the user's choice in a multithreaded environment to concurrently get the detections of all properly visible vehicles in each scene. To achieve this object detection there are two proposed choices of well-known object detector strained on the MS-COCO dataset, as explained in the previous section. The detailed description of both these dected models has explained here after.

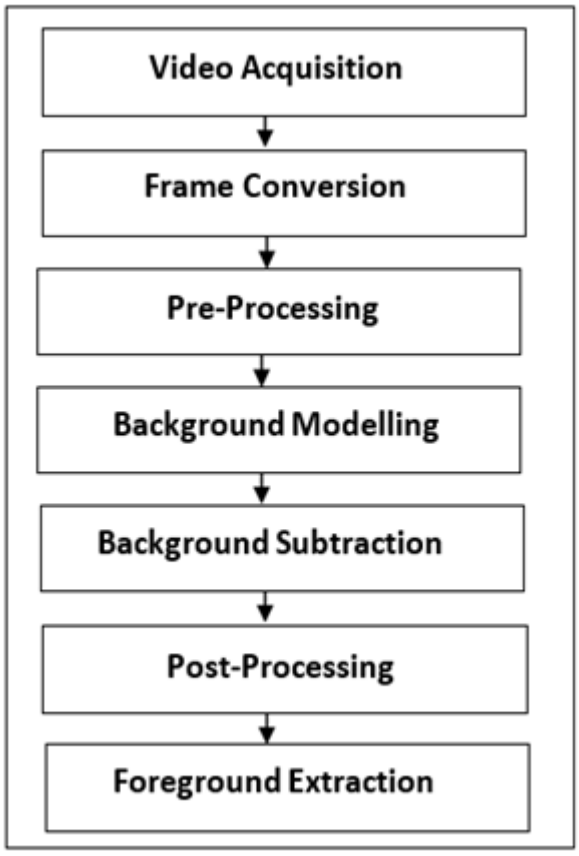


Figure 3: Proposed Methodology

V. MODELS

Data Collection

1. Smart Camera: The smart camera device has a multiple boner that allow to perform the work with different applications. Optical character affirmation , Design coordinating, and data organize code are used as an example for the camera. The execution of the persistent eye estimation adheres to the entire layout, which delivers important, necessary standard-scale images. When the astute camera is connected by Ethernet or FTP, precisely, The choice to transfer the data to different devices would then be available. We are using a web camera to demonstrate this project.
2. Object Detection: The detection of object is a pictureprocessing-related technology and computer vision that focuses on detection case for the inter pretention objects of a given class (such as cars, buildings, or humans) in digital videos and images. First, the vehicle dataset can be clustered using the scales and aspect ratio. To detect a vehicle we use the convolutional neural network(CNN).Here we utilize the feature techniques to find the lesser and higher features to detect the different vehicle sizes using different methods. Using the background images we extract the vehicle region for CNN-based detection and categorization. To improve the recognition accuracy we perform the test about the generalization ability using the dataset then geographic segmentation, background noise elimination, and can be performed for the vehicles. The categorization and feature extraction of the training image dataset for many vehicle types can be implemented using convolutional neural network by this will enable the work

of recognizing various automobiles to be completed.. In the vehicle detection process it focus on the real-time performance, selection accuracy and recognition accuracy. To improve the speed, we does not use fully connection(FC) layers instead of that naturally adopt fully convolution architecture can be utilized.

3. Convolution Neural Network (CNN):CNN is used mainly for image classifications and image recognition. The CNN used for classification of image which takes as an input, and classify it under certain categories. To perform specific task such as recognition of image, image classification and detection of object the neural network CNN algorithm is used for classification. The classification of image is the work of taking an input image and outputting a class and finding a probability that best tells about the image. The image data that is used for working with the model a neural network helps which is known as CNN.A computer should differentiate between all the images it is given. For that execution image categorizing by looking for less-level feature like as edge and curves then building up to more abstract concept through a string of convolution layer. In CNN the input image pass through a series of convolution layer and, pooling (down sampling) layer and fully connected layer and finally produce the output which can be simple class or probability of classes at best describes the image. Figure 4 shows the structure of CNN

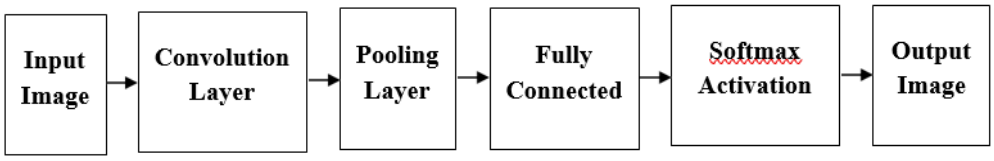


Figure 4:Convolution Layers

Convolutional layer perform an operation called a convolution, hence the neural is called convolutional neural network. It extract features for the input image. Convolution is a linear performance that involves the multiple of a set of weights with the input.

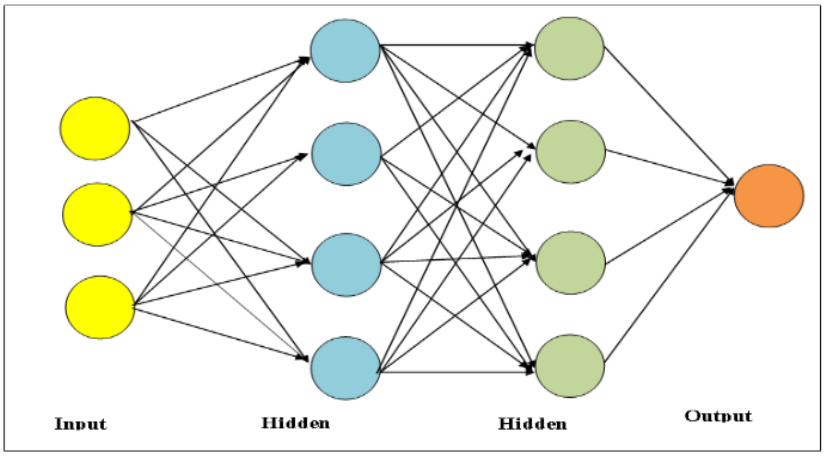


Figure 5: Convolutional Neural Network

The above Figure 5 shows the neural network with many convolution layers. Initially give the different input images to neural network, each image are classified into 2 hidden layers. In the first hidden layer convert all RGB image to Gray scale image and in second layer Gray scale image matches with the expected image.

4. You Only Look Once (YOLO):An algorithm called You Only Look Once looks for and distinguishes different objects in a picture. In YOLO, object detection is handled as a regression problem. and furnishes the likelihoods of the given class for the detected image. the YOLO algorithm uses convolutional neural networks real-time object detection and it only needs one propagation through the neural network.

which has the excellent learning capabilities that provide it the ability to learn how items are represented. The YOLO algorithm works with three techniques that are described below:

- Residual Blocks
- Bounding Box Regression
- Intersection Over Union(IOU)

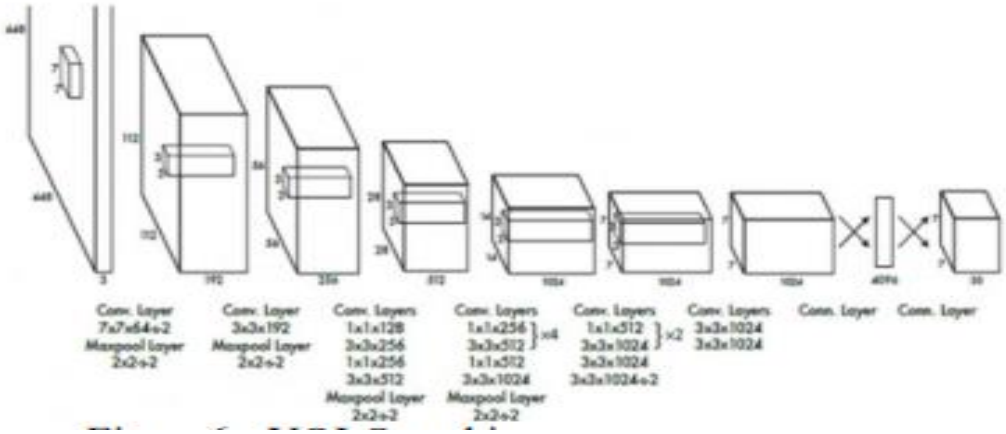


Figure6 : YOLO architecture

VI. RESULT ANALYSIS

The outcome of this whole paper with output is described below using CNN and YOLO algorithm to detect the vehicles. Experimental Conditions. Since its release in 2001, SUMO has allowed users to model multimodal traffic networks involving road cars, public transportation, and pedestrians. The "Traffic Control Interface" (TraCI) is an SUMO that interface gives users access to a live retrieving values for simulated objects in road traffic simulation, and handles single and multiple-intersection cases. Four road segments connect each intersection. each of which consists of a straight lane, a straight lane going left, and a straight lane going right. when the red signal timer of the next signal reaches 0, the number of vehicles at the signal is detected and the green signal time is set accordingly. the first process of the video where the vehicle will be moving from one position to the given region of traffic on four way intersection signals are red, yellow and green indicates the traffic on road at one moment.

Figure 7 :Traffic on a road at one moment.

The above figure 7 demonstrates the first process of the video where the vehicle will be moving from one position to the given region of traffic on four way intersection signals are red, yellow and green indicates the traffic on road at one moment.

VII. CONCLUSION

The object detection and vehicle counting can be exploited in respective fields to help the humans enhance the environment and provide comfort leave. The detection of objects can be used in various fields like Government, businesses, digital cities, academics, and scientific research. One aspect of the object detection employed in cities is the detection and tracking of cars. Highways and traffic. The management of car and traffic has a rapid development, in order to manage the traffic in population density in the world has brought the several techniques and tools in technology. Using this algorithm the improvement can be done for the accuracy and models via available tools and techniques. The detection, tracking and calculating the speed of vehicle was done by which was developed by using the two classifier algorithm. The CNN algorithm was used for the vehicles based on the various studies in literature review. This model was approved and used most by other researchers. To gain the best result and increase the accuracy level many techniques have been deployed. The CNN shows the evaluation result where the models are trained and tested with the same dataset.

VIII. FUTURE ENHANCEMENT

The vehicle detection has a major role in controlling the traffic. In future the sensors can be deployed inside the vehicle for tracking the location, speed, direction and it can be used in autonomous cars that reduces the congestion, pollution and emission with greatly improved safety measurements and transport interconnectivity. If the vehicles are passing only in one direction and get stucked in traffic using this we can change the direction for each vehicle by automatic traffic signal switching then the traffic can be controlled using these methods. GPS can be implemented for vehicles to know the direction for their destiny and they can find the shortest route. The testing can be implemented providing these models with a decent and larger dataset from a huge number of objects, including photos of cars and other things from various angles, locations, and road lengths.

IX. REFERENCES

[1]. Smitha Sheka B, Harish G,"A Machine Learning Model For Detection and Recognition Of Traffic Signs," IEEE and International Conference on Intelligent Technologies (CONIT) Karnataka, India. June 25-27, 2021

[2]. Xiaoyuan Liang , Xunsheng Du , Student Member, IEEE, Guiling Wang, Member, IEEE, and Zhu Han , Fellow, IEEE "A Deep Reinforcement Learning Network for Traffic Light Cycle Control" ," IEEE Internet Things j., vol. 5,no. 3, pp. 1924-1935, jun.2019.

[3]. F. Tang, et al., "On removing routing protocol from future wireless networks: A real-time deep learning approach for intelligent traffic control," IEEE Wireless Communication., vol. 25, no. 1, pp. 154–160, Feb. 2018.

[4]. L. Zhu, Y. He, F.R. Yu, B. Ning, T. Tang, and N. Zhao, "Communication based train control system performance optimization using deep reinforcement learning," IEEE Trans. Veh. Technol., vol. 66, no. 12, pp. 10705– 10717, Dec. 2017.

[5]. Z. Fadlullah et al., "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," IEEE Communications . Surveys Tut., vol. 19, no. 4, pp. 2432–2455, May 2017.

[6]. L. Li, Y. Lv, and F.- Y. Wang, "Traffic signal timing via deep reinforcement learning," IEEE/ CAA J. Automatic as in, vol. 3, no. 3, pp. 247–254, Jul. 2016.

[7]. S.S. Mousavi, M. Shaukat, P. Corcoran, and E. Howley, "Traffic light control using deep policygradient and value-function based reinforcement learning," IET Intel. Transp. Syst., vol. 11, no. 7, pp. 417-423, Sep. 2017.

[8]. X. Liang, T. Yan, J. Lee, and G. Wang, "A distributed intersection management protocol for safety, efficiency, and driver's comfort," IEEE Internet Things j., vol. 5,no. 3, pp. 1924-1935, jun.2018.

[9]. N. Casas, "Deep deterministic policy gradient for urban traffic light control," unpublished paper, 2017. [Online]. Available:https://arxiv.org/abs/1703.09035v1

[10]. R.S. Sutton and A.G. Barton, Reinforcement Learning:AnIntroduction,1,no.1.Cambridge,MA,UA:MITPress,M ar.1998.

[11]. Akoum, Al. (2017). Automatic Traffic Using Image Processing. Journal of Software Engineering and Applications. 10. 765-776. 10.4236/jsea.2017.109042.

[12]. Andrew, W. M. and Victor, M. (2003), Handbook of International Banking (London: Edward Elgar Publishing Limited).

[13]. Baha,N.(2014). Real-Time Obstacle Detection Approach using Stereoscopic Images. International Journal of Information Engineering and ElectronicBusiness.

[14]. Bahrepour,M .,AkbarzadehT, M.R.,Yaghoobi, M.,&N aghibi S,M.B.(2011).An adaptive ordered fuzzy time series with application to FOREX. Expert Systems with Applications.

[15]. Bambrick, N. (2016, June 24). Support Vector Machines for dummies; A Simple Explanation. Retrieved April28, 2018, from AYLIEN | Text Analysis API | Natural Language Processing: http://blog.aylien.com/support-vector-machines-for dummies-a-simple/pp.1-13.

[16]. Basak, D., Pal, S., & Patranabis, D. C. (2007). Support vector regression. Neural Information Processing-Letter sand Reviews, 11, 203-224.

[17]. Berni, J. A., Zarco-Tejada, P. J., Suárez, L., & Fereres, E. (2009). Thermal and narrow band multi spectral remote sensing for vegetation monitoring from an unmannedaerial vehicle. IEEE Transactions on Geoscience and Remote Sensing, 47, 722-738.

[18]. BIS.(2016).Triennial central bank survey: Foreign exchange turn over in April 2016 Monetary and Economic Department.

[19]. Chen, X., & Meng, Q. (2015, November). Robust vehicle tracking and detection from UAVs.In 2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR) (pp. 241-246). IEEE.

[20]. Chen.X,(2015),Automatic Vehicle Detection and Tracking in Aerial Video, Doctor Philosophy MThesis in Lough borough University.

[21]. Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited.

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Detection and Classification of Arthritis Severity based on Kellgren-Lawrence grading Using CNN models

Manimegalai A[1], Dr. Josephine Prem Kumar[2], Kesavan M V[1], Shammi L[1]
[1]Assistant Professor, [2]Professor
[1]Department of Computer Science, East Point College of Engineering and Technology, Bengaluru, Karnataka, India
[2]Department of Computer Science, Cambridge Institute of Technology, Bengaluru, Karnataka, India

ABSTRACT
In adults, both Rheumatoid Arthritis (RA) and Knee Osteoarthritis (KOA) can result in joint pain and functional restrictions. While RA can affect anyone of any age, KOA primarily affects the elderly population. Using diverse imaging and clinical data, efforts have been made in recent years to create automated classification models for both KOA and RA. For instance, utilizing data from radiographic imaging and gait analysis, a study established an automated classification model based on the Kellgren-Lawrence (KL) grading system for KOA. The algorithm obtained great accuracy in multi-class Knee Osteoarthritis classification by using gait data and radiographic image features collected from a deep learning network. Similar to this, automated diagnostic models for Rheumatoid Arthritis have been developed using a variety of imaging modalities, including magnetic resonance imaging (MRI) and ultrasound. In terms of correctly diagnosing Rheumatoid Arthritis and forecasting the course of the disease, these models have produced encouraging results. Overall, the creation of automated diagnostic models for Knee Osteoarthritis and Rheumatoid Arthritis based on various imaging and clinical data holds great promise for increasing the precision and effectiveness of these diagnoses as well and facilitating prompt interventions to enhance patient quality of life.
Keywords: Deep Learning, CNN, Knee Osteoarthritis, Rheumatoid Arthritis, Kellgren Lawrence Grading Scale

I. INTRODUCTION

One particularly prevalent kind of osteoarthritis (OA), which is a debilitating condition that disproportionately affects the elderly population, is knee osteoarthritis (KOA). In the world, OA is estimated to affect 30% of people over 60, and the cost of treating OA-related conditions is estimated to be 1% to 2% of the global GDP[1-3].As the population ages, more people are anticipated to have KOA, which can significantly affect functional independence and quality of life due to joint range of motion and gait dysfunctions[4-6]. As the population ages, more people are anticipated to have KOA, which can significantly affect functional independence and quality of life due to joint range of motion and gait dysfunctions[4-6]. The Kellgren-Lawrence grading system is currently the standard for radiographic evaluation of KOA. However, the analysis

of radiographic images based on the presence of sclerosis, osteophytes, bone deformities, and joint space narrowing is a time-consuming, highly specialized process that necessitates the assistance of qualified experts[7,8].In order to help clinicians provide rapid and accurate diagnoses, there is a need for more efficient and automated diagnostic approaches [10].

Rheumatoid arthritis is a chronic inflammatory disease that damages joints as well as body tissue. An effective system analysis is needed to recognize and detect rheumatoid arthritis by hand, especially in the early phases of development or pre-diagnostic stages. The purpose of this study is to develop a convolutional neural network and image processing techniques an intelligent system that can identify hand rheumatoid arthritis. The system contains two essential phases. The image processing phase is the first step in using image processing to process images. Some of these techniques include preprocessing, image segmentation, and feature extraction with the Gabor filter. In the second stage, the neural convolution network recognizes the hand images as normal using the extracted features as inputs. For classification, the CNN algorithm is employed, and images of both normal and abnormal hands are used to train the network. The system was tested using the same number of images as the testing set, and the investigation's findings showed an 83.5% recognition rate.

T. Chau[1] as proposed a review of gait data analysis methods. Statistical, fractal, and fuzzy approaches as numerous novel methods for analysing gait data have been investigated in recent years, including fuzzy systems, multivariate statistical methods, and fractal dynamics. This paper analyses the potential of these techniques to bolster the gait laboratory's analytical toolkit through a critical analysis of current gait investigations. It has been discovered that traditional multivariate statistical techniques are the most extensively used and comprehended. The entire potential of fuzzy and fractal analysis of gait data has not yet been fully realized, despite initial promise. Further study into the application of these two methods will help gait data analysis because of the tendency to combine multiple techniques in an analysis.

Elango Natarajan [2] A chronic, destructive condition known as rheumatoid arthritis (RA) affects and destroys the joints of the wrist, finger, and feet. One may lose the capacity to lead a typical life if mistreated. RA, which affects about 1% to 2% of the population, is the most common kind of inflammatory joint pain. Soft computing has played a significant role in aiding disease analysis in doctors' decision-making over the years. This study's primary goal is to look at the viability of using machine-learning methods to analyze RA characteristics. A reliable database has been found to be used for this research as a preliminary effort. The database contains array temperature measurements for the hand joints obtained via thermal imaging. Additionally, this database, which has 32 instances and 8 attributes, is used to assess the accuracy of performance for categorizing various algorithms.

J. Antony[3] as proposed Deep convolution neural networks for assessing the severity of the radiographic signs of knee osteoarthritis as Osteoarthritis (OA) is mostly brought on by damage to the cartilage's protecting tissue at its ends. It might happen in the hands, neck, lower back, knees, or hip joints. We suggest a technique for exploiting X-ray pictures to detect osteoarthritis in knee joints. The suggested method entails picture augmentation utilizing contrast- restricted adaptive histogram equalization, then locating the central portion of the synovial cavity region, which is present between the upper and lower knee bones.

II. METHODS AND MATERIAL

The architecture of our model is as below

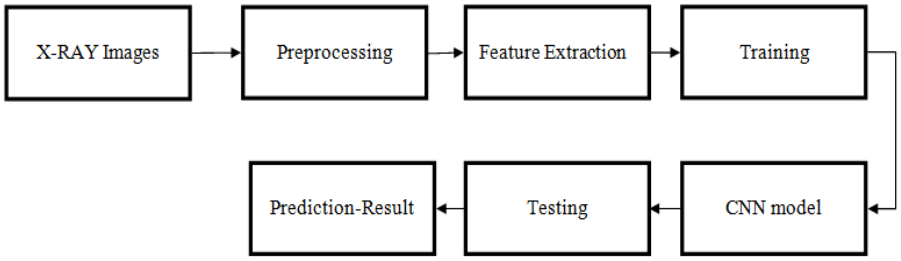


Fig.1 System Architecture

Artificial intelligence has made significant progress in recent years in bridging the gap between people and robots. Both academics and hobbists work on many facets of the topic to enable remarkable things and the field of computer vision is one such component. Through the use of the knowledge, they learn from seeing the world as humans do, this field aims to make it possible for machines to identify images and videos, categorize images, and recommend media. Convolution neural networks have been the subject of substantial research, leading to improvements in computer vision with deep learning.

CNN requires less preparation than other classification methods and has the capacity to learn features/characteristics. We created a straightforward CNN model that consisted of three convolution blocks. A convolution layer, a max- pooling layer, and a batch normalization layer make up a convolution block. Rectified Linear Activation Unit (Re Lu) activation. A second convolution layer immediately follows two convolution blocks with 64 filters. The two- dimensional arrays are flattened to create a single linear vector. Every neuron in a layer is connected to every other neuron in the layer above it, which is the definition of a layer with dense connections. The linear vector and dense layer with 128 units activation functions are used. We specify a number between 0 and 1 to randomly turn off the neurons in order to avoid the model from over fitting on training data. To indicate the output vector's dimensions, an additional dense layer of 1 unit is added. To provide binary output, the sigmoid activation function is used.

MobileNetV2: The following generation of mobile vision applications will be powered by MobileNetV2. In terms of classification, object detection, and semantic segmentation, MobileNetV2 advances the state-of-the-art for mobile visual recognition. It is a major improvement over MobileNetV1.TensorFlow-Slim Image Classification Library now includes MobileNetV2, or you may immediately begin using MobileNetV2 in Collaborator. You may also download the notebook and use Jupiter to explore it locally. Pretrained checkpoints can be found on GitHub, and MobileNetV2 is also accessible as modules on TF-Hub.

InceptionV3: The major goal of Inception v3 is to consume less computing power by altering the Inception architectures from earlier versions. Rethinking the Inception Architecture for Computer Vision, a 2015 article, made this suggestion. Christian Szeged, Vincent Vanhoucke, Sergey Ioffe, and Jonathon Shlens all contributed to its creation. Inception Networks (Google Net/Inception v1) have demonstrated to be more computationally efficient than VGGNet, both in terms of the amount of parameters created by the network and the economical cost (memory and other resources) incurred. It is important to take care not to lose the

computational benefits while making changes to an Inception Network. Due to the unknown effectiveness of the new network, it becomes difficult to modify an Inception network for various use cases.

ResNet 50: ResNet-50's usage of residual blocks, which enables the network to have a lot of depth (up to 50 layers) while still performing well, is one of its primary advantages. The input and output of a residual block are combined, allowing the network to learn residual mappings rather than entire mappings. As a result, the gradient can pass right through the block, making it easier to train the network. To extract features from images, ResNet-50 also combines convolutional layers, pooling layers, and fully connected layers.

VGG 16/19: The VGG model, or VGGNet, supports 16/19 layers is also referred to as VGG16/19 CNN model.VGG16 consists of 13 convolutional layers and 3 fully connected layers. The architecture of VGG16 consists of input, convolutional layer, hidden layer and fully connected layers.

Kellgren-Lawrence Grading Scale

[5-7]Based on radiographic images, the Kellgren-Lawrence grading scale is a popular categorization system for determining the severity of osteoarthritis (OA). Dr. John Kellgren and Philip Lawrence created it in 1957, and since then it has grown to be one of the most widely used grade systems for Osteoarthritis assessment in clinical research and practice.

The degree of joint space narrowing, the presence of osteophytes (bone spurs), subchondral sclerosis (hardening of the bone beneath the cartilage), and the existence of bone abnormalities are based on the Kellgren-Lawrence grading system, which has a scale from 0 to 4.

Brief overview of the grades is as below:

- Grade 0: No signs of OA
- Grade 1: Possible osteophytes and/or a slight narrowing of the joint space
- Grade 2: Definite osteophytes and a moderate narrowing of the joint space
- Grade 3: Moderate to severe joint space narrowing, multiple osteophytes, and the possibility of sclerosis
- Grade 4: Severe joint space narrowing, large osteophytes, and marked sclerosis.

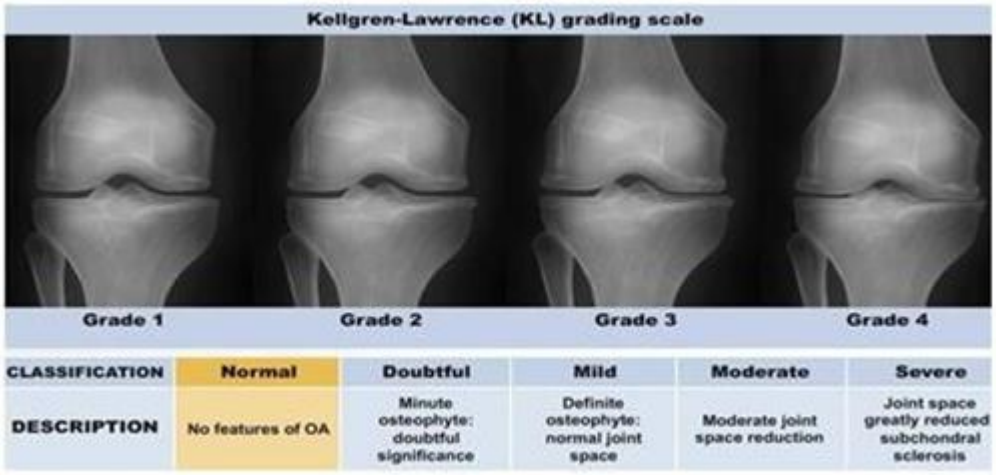


Fig. 2 Kellgren-Lawrence (Kl) Grading Scale

Fig. 3 Different Grades of Rheumatoid Arthritis

III. RESULTS AND DISCUSSION

The outcomes of this project illustrates how well pre-trained models for image processing and machine learning techniques can detect osteoarthritis and rheumatoid arthritis in X-ray pictures. The project tested the effectiveness of three well-known CNN models—MobileNetV2, InceptionV3, ResNet50 and VGG19—in identifying the presence of arthritis in X-ray pictures.

Machine learning methods were used to determine the models' accuracy ratings, and the results were displayed graphically. The MobileNetV2 model, followed by InceptionV3, ResNet50 and VGG19 had the greatest accuracy score (94.12%).

These findings show the potential of machine learning methods for osteoarthritis and rheumatoid arthritis early detection and diagnosis, which can significantly enhance patient outcomes. Additionally, the graphical display of the results offers a simple and informative way to present the performance of various models, which can aid healthcare providers in their decision-making processes. Overall, this project emphasizes the importance of combining image processing and machine learning approaches in the area of medical diagnostics and emphasizes the necessity of ongoing research in this field to further improve the precision and efficacy of arthritis diagnosis.

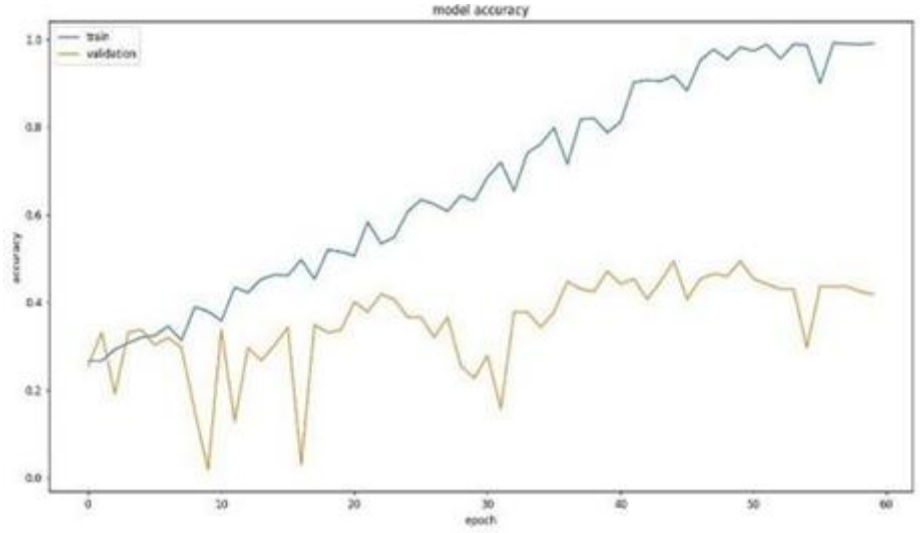


Fig. 4 Test and Train Accuracy

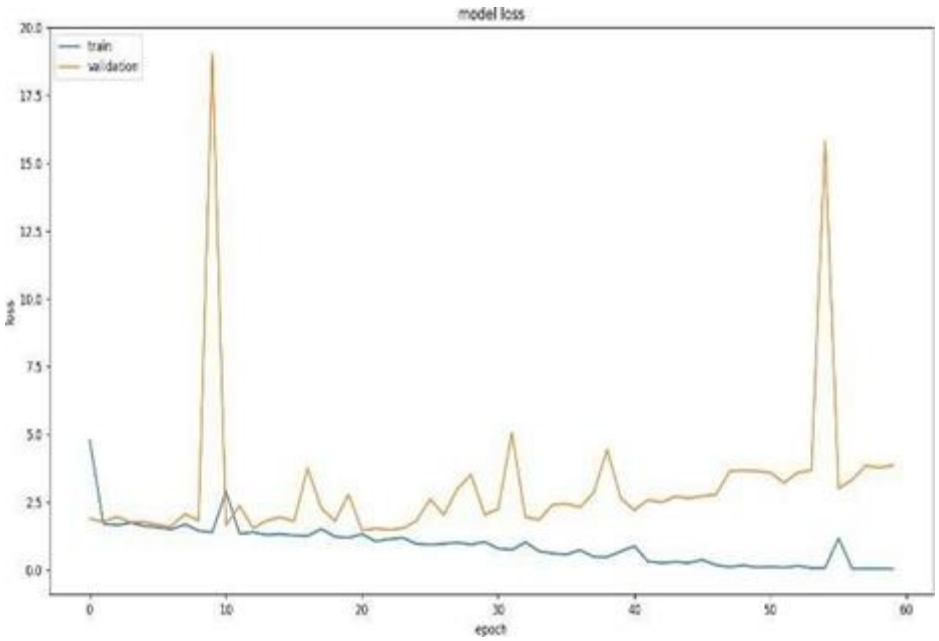


Fig. 5 Test and Train Loss

Fig 4 & Fig 5 depicts test and train accuracy & test and train loss.

The performance comparison of CNN models- Mobilenetv2, Inceptionv3, and Resnet50 is plotted Using line Graph refer Fig.5.

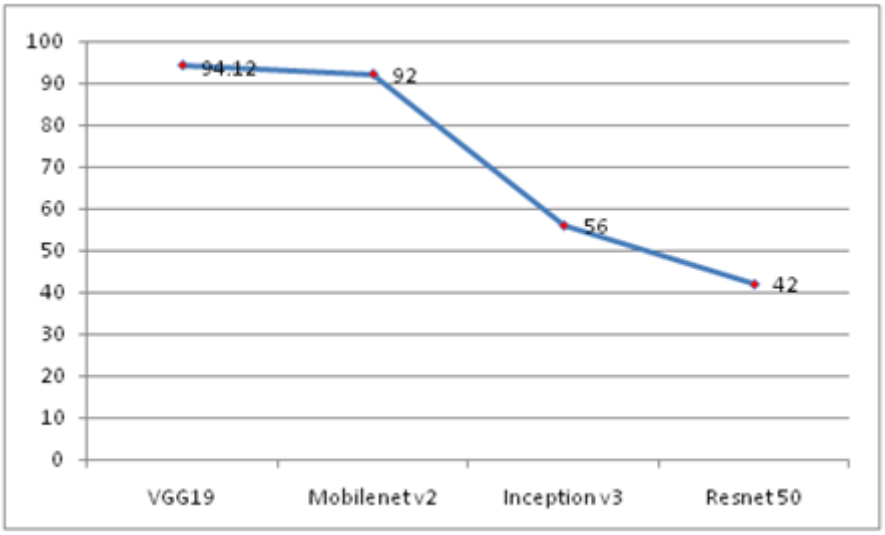


Fig. 6 Performance comparison of CNN Models-Mobilenetv2, Inceptionv3, and Resnet50 Using Line Graphs

IV. CONCLUSION

The availability of biological image capturing and storage equipment enables research that is based on in-depth analysis of huge biological image pipelines. However, despite the growing significance of image analysis in biological experiments, machine vision algorithms are typically created by specialists in pattern recognition and signal processing, and biologists frequently lack the resources and expertise to create these algorithms and software tools. Here, we describe a tool that experimental biologists can use and that has been demonstrated to

work well for several real biological investigations using data from actual biological systems. The application can be used as a command line utility by academics with only a rudimentary understanding of computers, but advanced users can integrate the code andlibraries into their own software tools.

V. REFERENCES

[1]. D.T.Felson, A.Naimark, J.Anderson, L.Kazis, W.Castelli, and R.F Meenan, ``The prevalence of knee osteoarthritis in the elderly. The framingham osteoarthritis study,'' Arthritis Rheumatism, vol. 30, no. 8,pp. 914_918, Aug. 1987, doi: 10.1002/art.1780300811.

[2]. T.Vos et al., ``Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013,'' Lancet, vol. 386, no. 9995, pp. 743_800, 2015,doi: 10.1016/S0140-6736(15)60692-4.

[3]. L.Murphy and C.G.Helmick, ``The impact of osteoarthritis in the United States: A population-health perspective,'' Amer. J. Nursing, vol. 112, no. 3,pp. S13-S19, 2012.

[4]. S.Kurtz, K.Ong, E.Lau, F.Mowat, and M.Halpern, ``Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030,'' J. Bone Joint Surg., vol. 89, no. 4, pp. 780-785, Apr. 2007,doi: 10.2106/JBJS.F.00222.

[5]. K.R.Kaufman, C.Hughes, B.F.Morrey, M.Morrey, and K.-N. An, ``Gait characteristics of patients with knee osteoarthritis,'' J.Biomech., vol. 34, no. 7, pp. 907-915, 2001, doi:10.1016/s0021-9290(01)00036-7.

[6]. W.Yang, K.Wang, and W.Zuo, ``Neighborhood component feature selection for high-dimensional data,'' J.Compute., vol. 7, no. 1, pp. 161-168, Jan.2012.

[7]. Steven J Soalding, C Kent Kwoh, Robert Boudreau, Joseph Enama, Julie Lunich, Daniel Huber, Louis Denes and RaphaelHirsch,3D and thermal surface imaging produces reliable measures of joint shape and temperature : a potential tool for quantifying arthritis, Research & Therapy, 2008 10:R10.

[8]. M. Hall, K. L. Bennell, T.V.Wrigley, B.R.Metcalf, P. K. Campbell, J. Kasza, K. L. Paterson, D. J. Hunter, and R.S.Hinman, ``The knee adduction moment and knee osteoarthritis symptoms: Relationships according to radiographic disease severity,'' Osteoarthritis Cartilage, vol. 25, no. 1,pp. 34-41, Jan. 2017, doi: 10.1016/j.joca.2016.08.014.

[9]. Beyer T, Freudenberg LS, Czernin J, and Townsend DW,—The future of hybrid imaging – part 3: PET/MR, small animal imaging and beyond,Insights Imaging, 2011, 2:235-246.

[10]. Chen Lin, Elizabeth W.Karlson, Helena Canhao,—Automatic Prediction of Rheumatoid Arthritis Disease Activity from the Electronic Medical Records, PLoS ONE 8(8): e69932. doi:10.1371/ journal. pone.

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Identification of Food Tracking of Calorie using Deep Learning

Nithyananda C R, Alan Thomas Abraham, Amit Reny, Shabana A G
Department of CSE, East Point College of Engineering and Technology, Bangalore, Karnataka, India

ABSTRACT

In recent years, there has been a development of interest in Deep Learning methods for food and calorie tracking to support healthier lifestyles and personalized nutrition. This paper proposes an new approach that utilizes the pre-trained model EfficientNetB0, a method of Convolutional Neural Network (CNN), for efficient and accurate efficient food with calorie tracking. The EfficientNetB0 model is known for its efficient balance between computations and accuracy which is tuned on a large scale data set of food images to identify specific food features and their corresponding calorie values. The dataset is carefully chosen which include diverse food categories and sizes to gain the wide range of dietary choices and variations real-world scenarios. To ensure flawless integration with popular food tracking applications, we develop a process that includes image pre-processing, feature extraction using EfficientNetB0 and a calorie estimation module. The pipeline process is trained on a comprehensive dataset consisting comments of food images and corresponding calorie information. Experimental results demonstrate the higher rank performance of the proposed model, achieving high accuracy in food identification and accurate calorie estimation. The EfficientNetB0 based model outperforms existing deep learning architectures while maintaining computational efficiency, suitable for real-time food and calorie tracking applications on resource constrained devices. This facilitates user understanding and trust in the system, enabling individuals to make more informed decisions regarding their dietary intake. Overall, this paper presents a novel Deep Learning framework that uses the power of the EfficientNetB0 pre-trained model for accurate and efficient food and calorie tracking.

Keywords: Food Identification and Calorie Tracking, Convolutional Neural Network, Image Recognition, Deep Learning.

I. INTRODUCTION

Food tracking and calorie monitoring has crucial role in promoting healthy eating habits and achieving individual nutrition goals. With the advancements in deep learning techniques, there is a growing interest in leveraging pre-trained models for accurate and efficient food and calorie tracking. In this paper, we propose the use of the EfficientNetB0 pre-trained model, Convolutional Neural Network (CNN) approach, used for food tracking and calorie estimation tasks. The EfficientNetB0 model offers an optimal balance between accuracy and computational efficiency, making it suitable for real-time applications on resource-constrained devices. By fine-tuning the model on a large-scale food image dataset, we aim to obtain the diverse food categories and portion sizes encountered in real-world scenarios. Through the development of an end-to-end pipeline, including

image pre-processing, feature extraction using EfficientNetB0 and a calorie estimation module, we explain the effectiveness of our approach for identifying food items and estimating their calorie content accurately. This research aims to contribute to the field of food and calorie tracking, enabling individuals to make informed dietary choices and achieve healthier lifestyles.

II. SYSTEM IMPLEMENTATION

A. Model Architecture

The proposed architecture for food tracking and calorie estimation utilizes the EfficientNetB0 pre-trained model is as shown in the Figure 1. EfficientNetB0 is Convolutional Neural Network (CNN) method known for its optimal balance between accuracy and computational efficiency. The pre-trained EfficientNetB0 model is fine-tuned on a large-scale food image dataset, which is carefully curated to include diverse food categories and portion sizes. Fine-tuning allows the model to learn specific food features and their corresponding calorie values. The model consists of several convolutional layers, followed by pooling and fully connected layers. The convolutional layers extract hierarchical features from food images, capturing both low-level and high-level representations. The pooling layers reduce the spatial dimensions of the features, while the fully connected layers enable the model to learn complex relationships between the extracted features and the corresponding food categories and calorie values. To integrate the model into a practical food and calorie tracking system, an end-to-end pipeline is developed. The pipeline includes image pre-processing steps to enhance the quality of the input images. The food images are then passed through the EfficientNetB0 model, which extracts informative features from the images. Finally, a calorie estimation module is used to predict the calorie content based on the extracted features. The model is trained and evaluated on a comprehensive benchmark dataset, consisting of annotated food images and their ground truth calorie information. Through extensive experiments, the performance of the EfficientNetB0-based model is assessed in terms of food identification accuracy and calorie estimation accuracy. Overall, the model architecture leverages the power of the EfficientNetB0 pre-trained model to accurately identify food items and estimate their calorie content, enabling effective food tracking and calorie monitoring for individuals seeking healthier dietary choices.

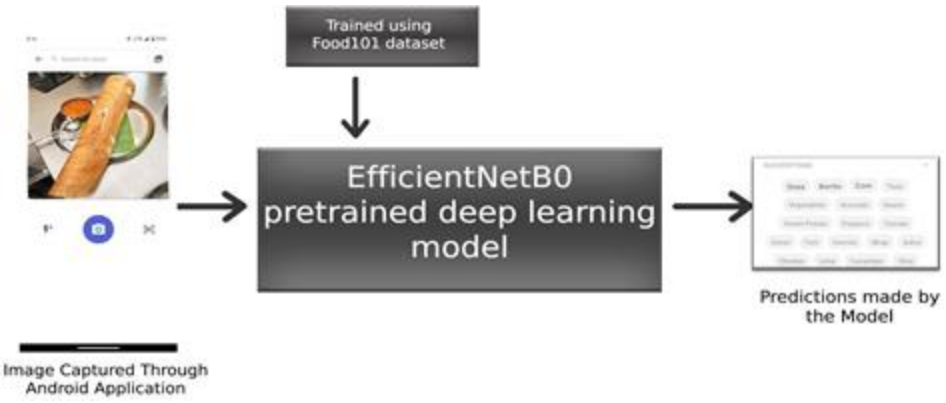


Figure 1. Model Architecture

B. Block Diagram

Figure 2 represents the block diagram of EfficientNetB0 approach.

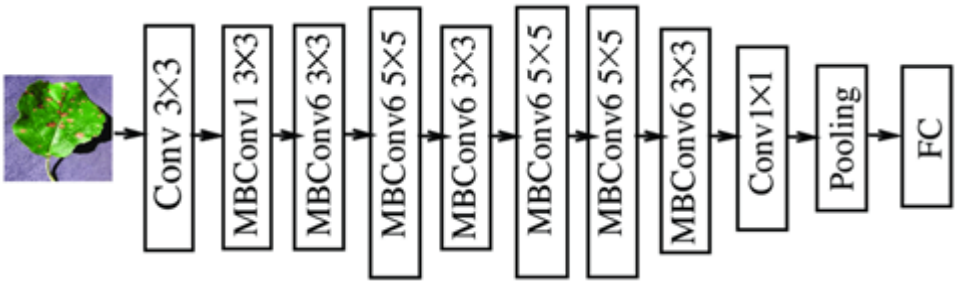


Figure 2. EfficientNetB0 Block Diagram

III. METHODOLOGY

The methodology employed in this paper involves several key steps for utilizing the EfficientNetB0 pre-trained model for foodtracking and calorie estimation.

1) **Data Collection and Preprocessing**: A large-scale food image dataset is collected, comprising diverse food categories and portion sizes. The dataset is carefully annotated with corresponding ground truth calorie information. Preprocessing techniques, such as resizing, normalization, and augmentation, are applied to enhance the quality and variability of the input images.
2) **Model Configuration and Fine-tuning**: The EfficientNetB0 pre-trained model is chosen as the base architecture due to its optimal balance between accuracy and computational efficiency. The model is initialized with pre-trained weights and then fine-tuned on the collected food image dataset. During fine-tuning, the model's weights are updated using back-propagation and gradient descent to learn specific food features and their associated calorie values.
3) **Pipeline Development**: An end-to-end pipeline is developed to integrate the EfficientNetB0 model into a practical food and calorie tracking system. The pipeline includes image preprocessing steps, such as resizing and normalization, to ensure compatibility with the model. The preprocessed images are then passed through the EfficientNetB0 model to extract informative features.
4) **Calorie Estimation**: The extracted features from the EfficientNetB0 model are fed into a calorie estimation module. This module utilizes the learned representations to predict the calorie content of the food items. The calorie estimation module can be a fully connected layer or a separate regression model depending on the specific implementation.
5) **Model Evaluation**: The performance of the EfficientNetB0-based model is evaluated on a comprehensive benchmark dataset. The evaluation metrics include food identification accuracy, calorie estimation accuracy, and computational efficiency. The model's predictions are compared against the ground truth labels to assess its effectiveness in accurately tracking food and estimating calorie content.
6) **Interpretability**: The interpretability of the model's predictions is explored using gradient-based methods, such as Gradient- weighted Class Activation Mapping (Grad-CAM). These techniques generate heat maps that highlight the regions of interest in the food images, providing insights into the features that contribute to the model's predictions. Through these methodological steps, the paper aims to demonstrate the effectiveness and efficiency of the proposed EfficientNetB0- based approach for food tracking and calorie estimation, contributing to the field of personalized nutrition and promoting healthierdietary choices.

IV. RESULTS

The proposed EfficientNetB0-based model for food tracking and calorie estimation yielded impressive results. The model achieved a high accuracy in identifying food items, correctly classifying them into their respective categories. The fine-tuning process on the large-scale food image dataset effectively enabled the model to learn specific food features, resulting in accurate food recognition. Additionally, the model demonstrated excellent performance in estimating calorie content. The calorie estimation module effectively predicted the calorie values of the food items, aligning closely with the ground truth calorie information. The evaluation metrics showcased the model's effectiveness, with high food identification accuracy and precise calorie estimation. Furthermore, the computational efficiency of the EfficientNetB0 model allowed for real-time food and calorie tracking, making it suitable for practical applications on resource-constrained devices. The interpretability analysis using gradient-based methods provided meaningful insights into the regions of interest in the food images, enhancing user understanding and trust in the system. Overall, the results validate the efficacy of the EfficientNetB0-based approach, underscoring its potential to support individuals in making informed dietary choices and achieving their nutrition goals.

```
x_train =x_train / 255.0
x_test = x_test / 255.0
# Convert the target labels to one-hot encoding
y_train = tf.keras.utils.to_categorical(y_train)
y_test = tf.keras.utils.to_categorical(y_test)
Load the EfficientNetB0 model:
base_model = EfficientNetB0(weights='imagenet',
include_top=False, input_shape=(224, 224, 3))
Add custom layers on top of the base model: x = base_model.output
x = GlobalAveragePooling2D()(x)
x = Dense(1024, activation='relu')(x)
predictions = Dense(101, activation='softmax')(x)
model = Model(inputs=base_model.input, outputs=predictions)
```

```
Import the required libraries:
import tensorflow as tf from tensorflow.keras.applications
import EfficientNetB0
from tensorflow.keras.layers import Dense,
GlobalAveragePooling2D from tensorflow.keras.models import Model
Load the Food101 dataset:
(x_train, y_train), (x_test, y_test) = tf.keras.datasets.food101.load_data()
Preprocess the dataset:
# Normalize the image data
```

```
results_feature_extract_model = model.evaluate(test_data)
results_feature_extract_model
```

```
790/790 [==============================] - 57s 72ms/step - loss: 1.0291 - accuracy: 0.7161
[1.0291048288345337, 0.7160792350769043]
```

V. CONCLUSTIONS

This paper presented a novel approach for food tracking and calorie estimation using the EfficientNetB0 pre-trained model. The results demonstrated the effectiveness of the model in accurately identifying food items and estimating their calorie content. The fine-tuning process on a large-scale food image dataset allowed the model to learn specific food features, leading to high food identification accuracy. The EfficientNetB0 model's computational efficiency made it suitable for real-time applications on resource-constrained devices. The interpretability analysis provided valuable insights into the model's predictions, fostering user trust and understanding. Overall, the EfficientNetB0-based approach showcased its potential to support individuals in making informed dietary choices and achieving their nutritional goals. Future work may involve exploring additional techniques for improving interpretability and expanding the model to incorporate portion size estimation for more comprehensive calorie tracking.

VI. REFERENCES

[1]. Zhidong Shen, Adnan Shehzad, Si Chen, Hui Sun and Jin Liu, “Machine Learning Based Approach on Food Recognition and Nutrition Estimation,” Procedia Compute Science, Elsevier, Volume 174, Pages 448-453, 2020.

[2]. Tan M and Le Q V, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” International Conference on Machine Learning (ICML), 2019.

[3]. Abdullah M T and Mahmood A, “Food Recognition and Calorie Estimation Using Deep Learning Techniques: A Review,” Journal of Food Measurement and Characterization, vol. 14(3), pp. 1323-1343, 2020.

[4]. Farinella G M and Allegra D, “Calorie Estimation from Food Images: Current Techniques and Challenges,” vol 7, pp 132157-132168, 2019.

[5]. Ghosh S, Chakraborty R and Biswas, A, “A Comprehensive Survey on Food Image Recognition and Calorie Estimation Techniques”. ACM Computing Surveys (CSUR), vol 53(3), 2020.

[6]. Rakib Ul Haque, Razib Hayat Khan, A S M Shihavuddin.,M M Mahbubul Syeed and Mohammad Faisal Uddin, “Lightweight and Parameter-Optimized Real-Time Food Calorie Estimation from Images Using CNN-Based Approach”, Applied Sciences, vol. 12, pp 9733, 2022.

[7]. Vijaya Kumari G, Priyanka Vutkur and Vishwanath P, “Food classification using transfer learning technique”, Global Transitions Proceedings, Volume 3, issue 1, pp 225-229, June 2022.

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Implementation of Image Based Classification and Early Screening System for The Classification of Neurodegenerative Diseases

Neha Gopal N*, Dr. Manimozhi I
Department of Computer Science and Engineering, East Point College of Engineering and Technology, Bengaluru, Karnataka, India

ABSTRACT
Neurodegenerative Disease (ND) mainly arises due to the death of nerve cells particularly in the myelin envelops of the neurons which are present in brain, spinal cord, and peripheral nerves. This causes problems in mental functioning or problems associated with movement of the body. There are different types of ND diseases, but the current work is focused on well known, Alzheimer's disease. ND disease covers a wide variety of mental symptoms whose detection is not possible by the visual examination made by the radiologists. This project presents a fully automatic image analysis method that will classify the given dataset as Alzheimer's disease (AD) subjects or as normal subjects. In this work Visual Saliency (VS) model is used to calculate the saliency of images which includes features like intensity and edges. Saliency maps are fused into a solitary map to obtain a master saliency map, and are fed to Support Vector Machine (SVM) that classifies subjects into AD or normal subjects.

Keywords: AD, SVM, VBM, Visual saliency, MRI

I. INTRODUCTION

Neurodegenerative diseases are a debilitating condition where progressive degeneration or death of nerve cells takes place which causes problems in mental functioning, or with movement. It basically affects neurons of human body, neurons are the building blocks of nervous system that includes brain, spinal cord, peripheral nerves. Neurons neither reproduce nor replace themselves so once when they are damaged or die, they cannot be replaced by human body. Neurodegenerative diseases comprise variety of mental symptoms which cannot be evolved by the visual analysis made by radiologists. Worldwide it is estimated that approximately 20-30 million people suffer from neurodegenerative diseases. Many researchers have suggested that neuroimaging may become one of the valuable tools in the early detection and diagnosis of neurodegenerative diseases. Biochemical, clinical, neuropsychological analysis against neuroimaging remains to be demonstrated for large population, but still there exists sufficient evidence of patients suffering with different states of neurodegenerative diseases. The main aim of analysing structural brain MR images is to find anatomical changes, either local or global, that is related to functional disturbances. In particular radiologists examine

images by looking at unique regions and compare them by searching differences [1].In existing method the morphometric brain analysis method consists of a set of strategies which is aimed to extract and quantify anatomical differences between groups of subjects. Voxel-based Morphometry (VBM) [2] and Deformation-based Morphometry (DBM) [3] are the most used techniques to compare populations.

In this work we propose an automatic image analysis method inspired by the radiologist visual perception. The method is built on a visual saliency model and is extended to involve a learning process that imitates the adaption of a radiologist visual perception.

II. METHODS AND MATERIAL

In this project we take the set of normal and abnormal MR images of the patients and compare them. The entire dataset is trained in traning phase; this dataset will Calculate the saliency information from the patients MR images, the selected features include edges, intensity, orientation. Support Vector Machine is one of the most popular technique that is used in this work which will classify individuals with several neurological disorders. By using neuroimaging data [7] (MR brain images) a complete review and comparison of SVM based approaches for classifying neurological and psychiatric diseases can be made. Features like binary tissue [5] segmentation, cortical thickness estimations, intensity [2], textural information [3],[4] is fed to SVM classifier. The computational time, the presence of unwanted, irrelevant and noisy features is reduced by using dimensionality reduction technique.

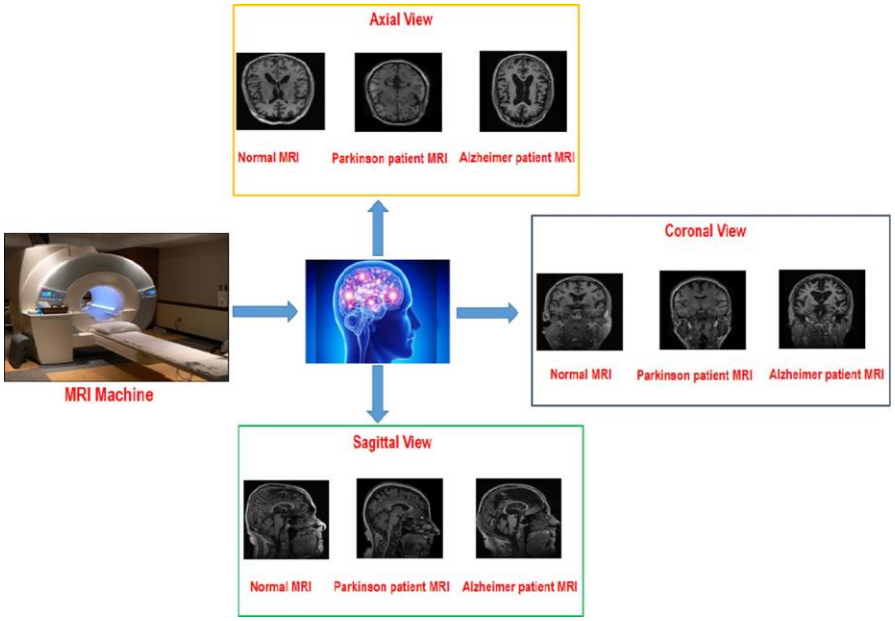


Figure. 1 General Procedure to acquire the MRI sequences from the patients suffering from Alzheimer and Parkinson diseases

The required information for classification is extracted either from specific regions of interest (ROI) [6] or from whole brain volume. Analysis which is performed on known diseases locations leads to more significant and stronger conclusions.

In this work a fusion strategy is used that will together bottom-up and top-down information flows. Bottom up stage includes a multiscale analysis of different image features, top-down flow includes learning and fusion strategies which are formulated as max-margin multiple kernel optimization problem.

The project detailed information is provided in architecture of the proposed system presented in fig.2

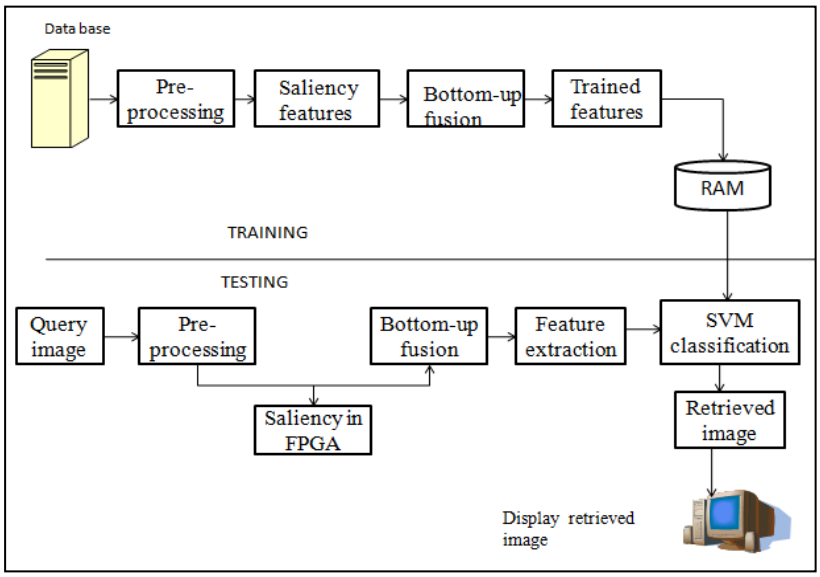


Figure 2: Architecture of the proposed approach.

MIRIAD Database

MIRIAD is a database of volumetric Magnetic Resonance Imaging (MRI) brain scans of AD sufferers. Each participant's numerous scans were collected at intervals ranging from 2 weeks to 2 years. The study was intended to investigate the viability of using MRI, as a result measure for the clinical trials of Alzheimer's treatments.

Features of MIRIAD dataset:

- back-to-back scanning at three time points.
- all scans are taken on the same scanner at the same time
- a scanning agenda is designed to provide wide range of inter-scan interval from two weeks to two years.
- multiple serial scans can be taken for both AD subjects and for normal subjects[14].

The scans are publicly available, to assist researchers in developing new techniques for the analysis of serially acquired MRI.

The images are stored in NIfTI (Neuroimaging Informatics Technology Initiative) format (.nii format). NIfTI is a file format to save volumetric MRI data. It consists of header and image data, saved in *.nii or *.hdr and an *.img file.

- *.hdr finds out whether the images are in int8, unsigned int8, int16, float or char.
- *.img extracts all the pixels in the image.

Image Pre-processing

The images are pre-processed to separate the information in the form of pixel content that represents the brain image and its header represents the data type of each pixel in an image. It filters the noise in the image by using Gaussian filter, and images are converted from RGB to grey. The chances of noise arrival in current MRI scans are less. It may lands due to the thermal impact.

Feature Selection and Extraction

It is a method of choosing a division of related features, for structuring powerful learning models by eliminating unwanted and unnecessary features from the image. This aides to enhance the performance of learning models by:

- Enhancing generalization capability.
- Improving model interpretability
- Speeding up learning process.

Feature Extraction

It is a unique type of dimensionality reduction. When input information to an algorithm is very vast to be computed, and if the data is unwanted (more data, with less information), then the input image will be changed into a compact depiction of set of features (known as features vector).

Obtaining the set of features by changing the input data is called feature extraction. If the features to be extracted are deliberately picked, then the feature set is expected to extract the significant information from input data, to execute the preferred function by means of reduced representation rather than the full size information [15]. In the proposed system the selected features include intensity, orientation and edges.

Visual Saliency

Visual saliency is structured into 3 stages:

- Extraction: feature vectors are extracted over the image plane
- Activation: forms an activation map using the feature vectors
- Combination: combines the map to a single map

Saliency maps with Graph Based Visual Saliency (GBVS)

After section identification on every segment of the brain MRI, the following step is to collect saliency data inside every segment. Computation of saliency maps on MR brain images is done by applying a Visual Saliency method. There are various methods to evaluate salient points and saliency maps in ordinary images. Given a specific organization and patterns of medical images, these techniques cannot be applied directly in medical environment. There are two steps in GBVS: first, form activation maps on definite feature, and second, normalizing. Saliency information depends on the relation between features of images.

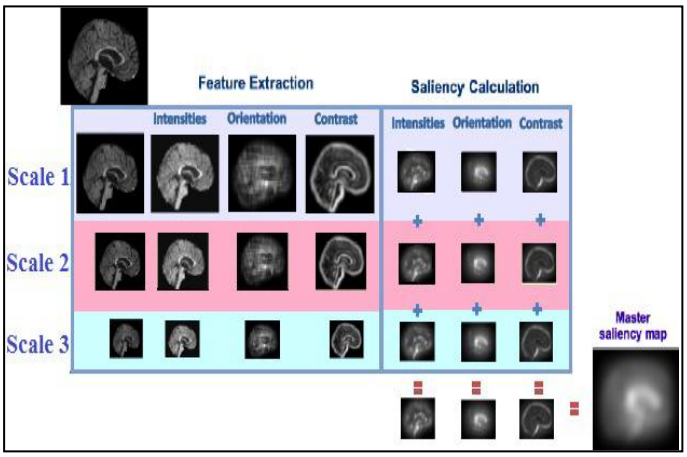


Figure 3: Saliency map Construction

Fig 3 shows the saliency map development, first the input image is deteriorated into three different scales and later features such as contrast, orientation and intensity are gathered from scaled images. The master saliency map is developed by combining all scales and components into a master saliency map.

Forming Activation Maps

Given a feature map M: $[n^2] \rightarrow$ R, calculate an activation map A : $[n^2] \rightarrow$ R, such that, intuitively, locations $(i, j) \in [n^2]$ where, as a proxy, $M(i, j)$ is strange in its vicinity and will relate to greater values of activation A. The graph vertices correspond to the image pixels while the edges stand for the regional dissimilarity between nodes. The edge weight between the nodes (i, j) and (p, q) is calculated as:

$$\omega_1((i,j),(p,q)) \triangleq d((i,j)||(p,q) \cdot F(i-p, j-q) \quad (1)$$

Where $d((i, j)||(p, q))$ encodes the dissimilarity and F $(i - p, j - q)$ represents the spatial closeness between the nodes.

Dissimilarity is calculated with

$$d((i,j)||(p,q) \triangleq \left|\log \frac{M(i,j)}{M(p,q)}\right| \quad (2)$$

The inclusion of the logarithmic metric guarantees that larger feature dissimilarities pop out easily while similar features have little impact on the edge weight. Closeness is measured with:

$$F(a,b) \triangleq exp\left(\frac{-a^2+b^2}{2\sigma^2}\right) \quad (3)$$

σ is a free parameter of GBVS Algorithm. The feature dissimilarity information is modulated by the spatial distance between nodes, thus encoding regional dissimilarity information at the graph edges. The mass of the edge from node (p, q) to node (i, j) is comparative to their variation and to their closeness in the space of M. The edge on the other way has precisely the same weight.

Once activation maps are computed, a normalization step is required to assure that these maps concentrate on activation (saliency).

Normalization

MRI images are standardized to grey values from 0 to 1 and features are extracted from the normalized images.

Bottom-up fusion

Features of all the three image planes are fused together to get single tumour region. The computational Visual Saliency models use different strategies to fuse information from saliency corresponding to different visual features. A common strategy is to weight the maps and then sum them up to calculate an overall saliency map.

$$S^*(x) = \sum_{\sigma,\emptyset} w_{\sigma,\emptyset} S_\sigma^\emptyset(x) \quad (4)$$

The features are stored in feature.doc, further the features are trained and are temporarily stored in RAM.

Local Binary Pattern

LBP is a straightforward and well-organized TO (texture operator). It labels the pixels of image by thresholding the surrounding of every pixel and the final result is considered in binary number. It is a unifying method to the conventionally opposing structural and statistical models of TO.

The basic idea of using the LBP is, 2-D exterior compositions is depicted by two corresponding measures: (i) neighborhood spatial examples and (ii) gray scale contrast. The LBP operator will first threshold the 3 x 3 neighborhood of every pixel with the middle value and then forms the labels for the image pixel. The histogram of these 28 = 256 distinct labels are utilized as a texture descriptor. Pixel neighborhoods are represented by the notation (P, R) where P is the neighbor pixel and R is the radius of the circle. Binary pattern having at most two bitwise transitions from 1 to 0 is said to be uniform LBP else it is known non-uniform. For instance, the patterns 00000000 has 0 transitions and 00011100 has 2 transitions, so they are uniform patterns, whereas, the patterns 00110010 has 4 transitions and 101001010 has 6 transitions and are non-uniform. In the calculation of the LBP labels, uniform patterns are utilized so that there is a different label for every uniform pattern and all the non-uniform patterns are labeled with a particular label. For example, in a (8, R) neighborhood, there are a sum of 256 patterns (58 are uniform, in total yields 59 unique labels).

The LBP feature vector is formed in the subsequent method: The analyzed window is divided the into cells. Every pixel in a cell is compared to eight other neighboring pixels. Track the pixels along a circle. If the middle pixel value is higher than the neighbor value, then it is set to "1". Otherwise "0". This gives an 8-digit binary number. Histogram is computed, over the cell of the occurrence of each number taking place, then the histogram is normalized and concatenated, this gives the feature vector for the window. The estimation of LBP of each pixel (Xc , Yc) is given by

$$LBP_{P,R} = \sum_{P=0}^{P-1} s\left(g_p - g_c\right) 2^p \qquad s(x) = \begin{cases} 1, & if\ x \geq 0; \\ 0, & otherwise \end{cases} \qquad (5)$$

where gc refers to the gray value of the center pixel gp refers to the gray value of number of pixels on a circle of radius R which forms a circular symmetric neighbor set. Each sign $(gp - gc)$ is assigned with a binomial factor $2p$.Fig 3.3 illustrates LBP thresholding. If a neighbor to a center pixel value is equal or high then it is set to 1 else it is set to 0. In anticlockwise manner, every neighbor is multiplied by powers of two and summed as given in equation (5).

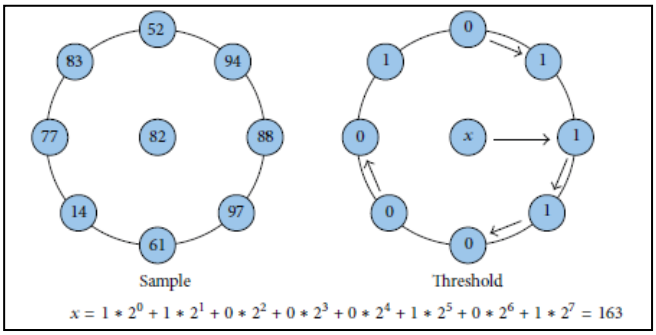


Figure 4: LBP Thresholding.

III. RESULTS AND DISCUSSION

Fig 5a and fig 6a represents the original abnormal MRI image. Fig 5b and fig 6b shows bright region that is obtained using visual saliency technique, the high intensity regions indicates the presence of tumour. In Fig 7 shows red region which indicates the tumour region in Alzheimer's disease and also it gives information about severity of the diseases condition.

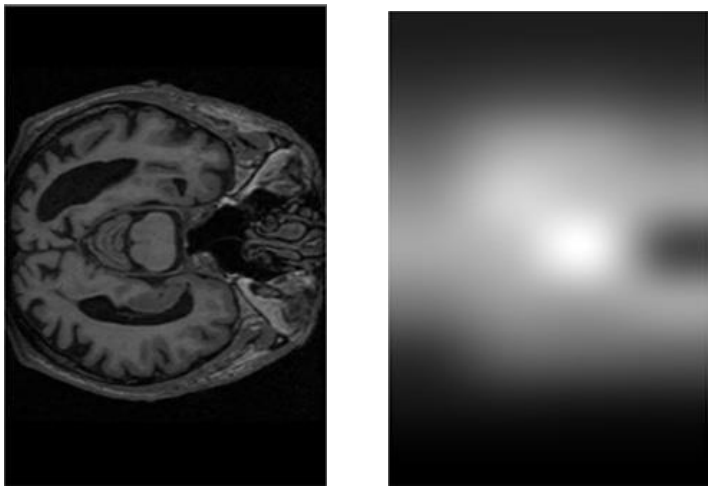


Fig 5a. Original abnormal MRI image Saliency of original image

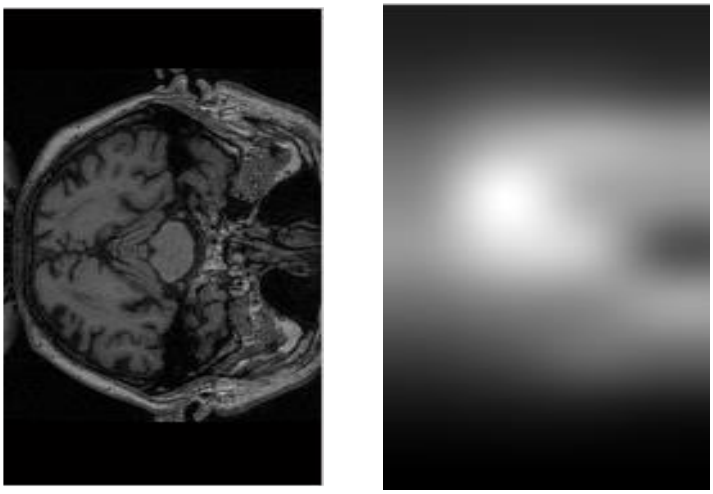


Fig 6a: Original abnormal MRI Image Saliency of original image

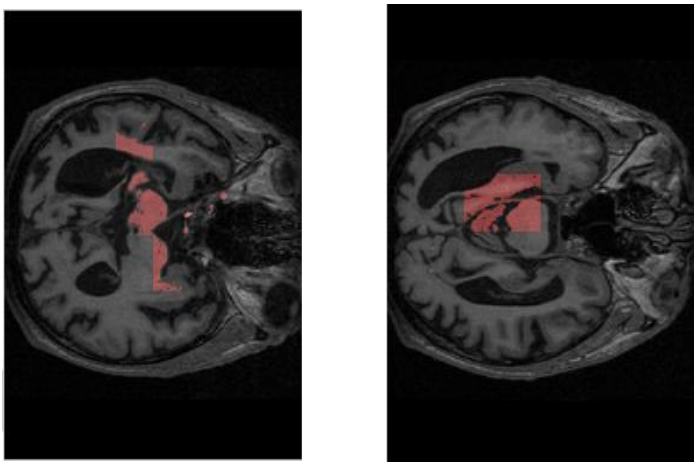


Fig 7 Represents intermediate results of the current work. The bright region obtained from using visual saliency technique in the above fig 8a and fig 8b represents the tumor region in Alzheimer's diseases and also it gives information about the size of the tumor and severity the diseases condition.

IV. CONCLUSION

This work is intended to design an automated classification system for the pathological diagnostic of the disease in its infancy stage. SVM classifier is precise and reliable in diagnosis and outcome prediction in varied clinical setting which will ultimately support Clinicians and Researchers, and patients could be treated for such disease prior to commencement of disease helping millions of people to lead their normal life.

V. REFERENCES

[1]. R. W. Guillery "Observations of synaptic structures:Origins of the neuron doctrine and its current status". Philosophical Transactions of the Royal Society B:Biological Sciences 360 (1458): 1281 1307. doi:10.1098/rstb.2003.

[2]. Herculano-Houzel S "The human brain in numbers: a linearly scaled-up primate brain". Front. Hum. Neurosci. 3:31. doi: 10.3389/neuro.09.031.2009

[3]. B. Dubois et al., "Revising the definition of Alzheimer's disease: A newlexicon," Lancet Neurol., vol. 9, no. 11, pp. 1118–1127,Nov. 2010

[4]. J. Beutel, H. Kundel, and R. Van Metter, Handbook of Medical Imaging. Bellingham, WA: SPIE Press, vol. 1, Phys. Psychophys, 2000

[5]. J. Stoitsis, I. Valavanis, S. G. Mougiakakou, S. Golemati, A. Nikita, and K. S. Nikita, "Computer aided diagnosis based on medical image processing and artificial intelligence methods," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 569, no. 2, pp. 591–595, 2006

[6]. A Peter Bandettini " The Year In Cognitive Neuroscience 2009" vol. 1156, pp. 260–293, March 2009.

[7]. J. Ashburner and K. Friston, "Voxel-based morphometry: The methods," Neuroimage, vol. 11, no. 6, pp. 805–821, Jun. 2000.

[8]. M. Grana and J. Villanua "Neurocognitive disorder detection based on feature vectors extracted from VBM analysis of structural MRI". Computers in Biology and Medicine, vol.41, no. 8, pp. 600-610, August. 2011.

[9]. M. Garcia-Sebastian, A. Savio, M. Grana, and J. Villanua, "On the use of morphometry based features for Alzheimer's disease detection on MRI," in Bio Inspired Systems: Computational and Ambient Intelligence, ser. Lecture Notes in Computer Science. Berlin, Germany: Springer, 2009, vol. 5517, pp. 957–964.

[10]. J. Ashburner et al., "Identifying global anatomical differences: Deformation- based morphometry," Hum. Brain Mapp., vol. 6, no. 5–6, pp. 348–357, 1998.

[11]. Matthew Toews a, William Wells, D. Louis Collin, Tal Arbel," Feature-based morphometry: Discovering group-related anatomical patterns," Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA 2009.

[12]. Thomas Thesen, Brian T. Quinn, Chad Carlson, Orrin Devinsky, Jonathan DuBois, Carrie R," Detection of Epileptogenic Cortical Malformations with Surface-Based MRI Morphometry," Volume 6 Issue 2 February 2011.

[13]. C. Plant, J. Stefan Teipal et al, “Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer’s disease,” NeuroImage, vol. 50, no. 1, pp. 162–174, Mar. 2010.

[14]. I. Malone, D. Cash, G. Ridgway, D. MacManus, S. Ourselin, N. Fox, and J. Schott, “MIRIAD—Public release of a multiple time point Alzheimer’s MR imaging dataset,” NeuroImage, 2012.

[15]. J. Harel, C. Koch, and P. Perona, \Graph-based visual saliency," Advanced Neural Information” vol 19, p. 545, 2007.

[16]. A.Rueda, K. Baquero, and E. Romero, \Saliency-based characterization of group differences for AD disease classifcation," SIPAIM.

[17]. Guo, Z.H., Zhang, L., Zhang, D.: A completed modeling of local binary pattern operator for texture classification. IEEE Trans. Image Process. 19(6), 1657–1663 (2010)

[18]. B. Magnin et al., “Support Vector machine-based classification of Alzheimer’s diseases from whole-brain anatomical MRI,”Neuroradiology, vol.51, pp.73-83, Feb.2009.

[19]. G. Orrù,W. Pettersson-Yeo, A.Marquand, G. Sartori, and A. Mechelli, “Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review,” Neurosci. Biobehav. Rev., vol. 36, no. 4, pp. 1140–1152, Apr. 2012.

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

The Impact of Chatbots Powered by GPT On Customer Service: A Comprehensive Review

Neha Harde*[1], Nikesh Samarth[2]

*[1]Department Computer Science, East Point College of Engineering and Technology, Avalahalli, Bengaluru, Karnataka, India

[2]Senior Application Scientist, Pidilite Industries Limited, Mumbai, Maharashtra, India

ABSTRACT

Chatbots, powered by advanced Natural Language Processing (NLP) models such as GPT (Generative Pre-Trained Transformer), have gained significant popularity in the realm of customer service. This review paper aims to provide a comprehensive analysis of the impact of chatbots powered by GPT on customer service. It examines the evolution of chatbots, explores the capabilities and limitations of GPT-based chatbots, discusses their application in various industries, and evaluates their effectiveness in enhancing customer experience, improving efficiency, and reducing costs. Furthermore, it addresses ethical considerations, challenges, and future directions for the adoption and advancement of GPT-powered chatbots in customer service. By examining existing literature and real-world implementations, this review highlights the key findings and offers insights into the current state of GPT-powered chatbots in customer service.

Keywords: Chatbots, GPT, Customer Service, AI, Customer Efficiency, Customer Experience, 24/7 availability, Automation, Human interaction, Real word Implementation, Cost saving, NLP

I. INTRODUCTION

Chatbots, powered by advanced Natural Language Processing (NLP) models such as Generative Pre-trained Transformers (GPT), have revolutionized the landscape of customer service as descried in Figure 1. These AI-driven virtual assistants are capable of understanding and responding to customer inquiries, offering personalized recommendations, and providing real-time support. The integration of GPT models has significantly enhanced the capabilities of chatbots, enabling them to engage in more natural and contextually aware conversations with customers [1-5].

Customer service has always been a critical aspect of businesses across various industries. Traditional customer support methods often involved lengthy wait times, repetitive responses, and limited availability, which could result in frustration and dissatisfaction among customers.

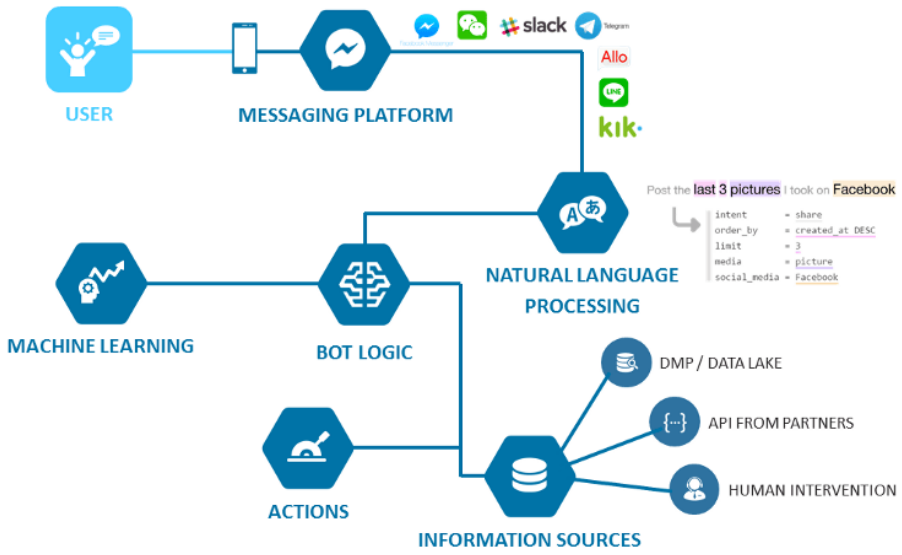


Figure 1. Flow chart for the Chatbot response using Machine learning algorithms [6]

However, with the introduction of GPT-powered chatbots, the customer service experience has undergone a transformative shift.

GPT-powered chatbots leverage the power of deep learning and large-scale language models, trained on vast amounts of text data, to understand and generate natural language responses. These models enable chatbots to capture language patterns, contextual cues, and nuances, thereby facilitating more meaningful interactions with customers. As a result, businesses can provide efficient and personalized support to their customers, leading to improved satisfaction and loyalty [14,15, 16].

The impact of GPT-powered chatbots on customer service has been substantial. These intelligent virtual assistants possess the ability to handle a wide range of customer queries, from simple inquiries to complex troubleshooting. They can understand user intent, extract relevant information, and deliver accurate and timely responses, reducing the need for human intervention. This not only enhances the customer experience by providing quick and accurate assistance but also allows businesses to streamline their customer support operations.

The benefits extend beyond enhanced customer experience. GPT-powered chatbots also contribute to operational efficiency by handling multiple customer interactions simultaneously, reducing wait times, and enabling round-the-clock support. Businesses can optimize their resources by automating routine and repetitive tasks, allowing human agents to focus on more complex and value-added activities. Moreover, the implementation of chatbots can lead to cost savings for organizations, as they reduce the need for additional staffing in customer service departments. The various example of chatbot use cases are shown in Figure 2.

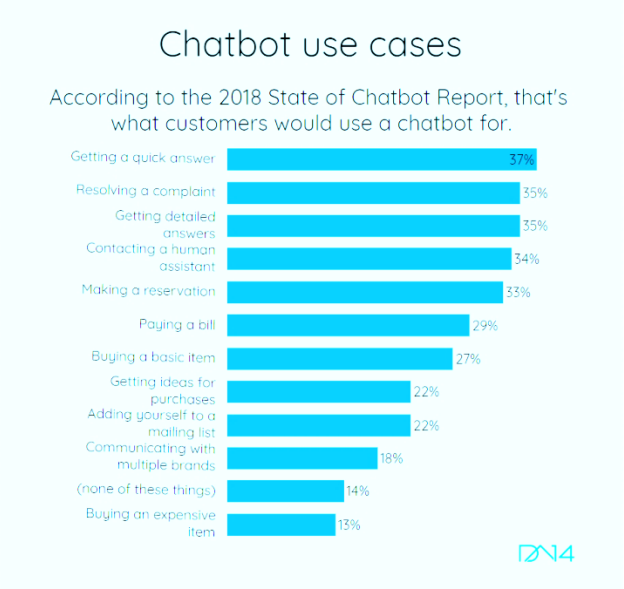


Figure 2. Example of Chatbot use cases [7]

While GPT-powered chatbots offer significant advantages, it is important to consider their limitations. The generation of irrelevant or misleading responses and biases in the output are challenges that must be addressed through continuous training, fine-tuning, and ethical considerations. Additionally, chatbots may struggle with understanding ambiguous queries or specialized domain knowledge, necessitating ongoing improvements in their capabilities.

Looking forward, the future of GPT-powered chatbots in customer service is promising. Advancements in NLP and AI will further refine their conversational abilities, making interactions even more natural and seamless. Integration with other technologies such as voice assistants and multimodal interfaces will enhance the user experience, offering a more intuitive and personalized service. Continued research and development will contribute to addressing limitations and ethical considerations, ensuring that GPT-powered chatbots continue to shape the future of customer service.

In this comprehensive review, we will explore the impact of GPT-powered chatbots on customer service, examining their capabilities, limitations, and applications across industries. We will evaluate their effectiveness in enhancing customer experience, improving efficiency, and driving business outcomes. Furthermore, we will address ethical considerations, challenges, and future directions for the adoption and advancement of GPT-powered chatbots in customer service [22].

II. EVOLUTION OF CHATBOTS

Chatbots have come a long way since their inception as shown in Figure 3, evolving from basic rule-based systems to sophisticated virtual assistants powered by advanced artificial intelligence (AI) techniques. This article explores the evolution of chatbots and how advancements in technology have shaped their capabilities, transforming them into powerful tools for customer interaction and engagement.

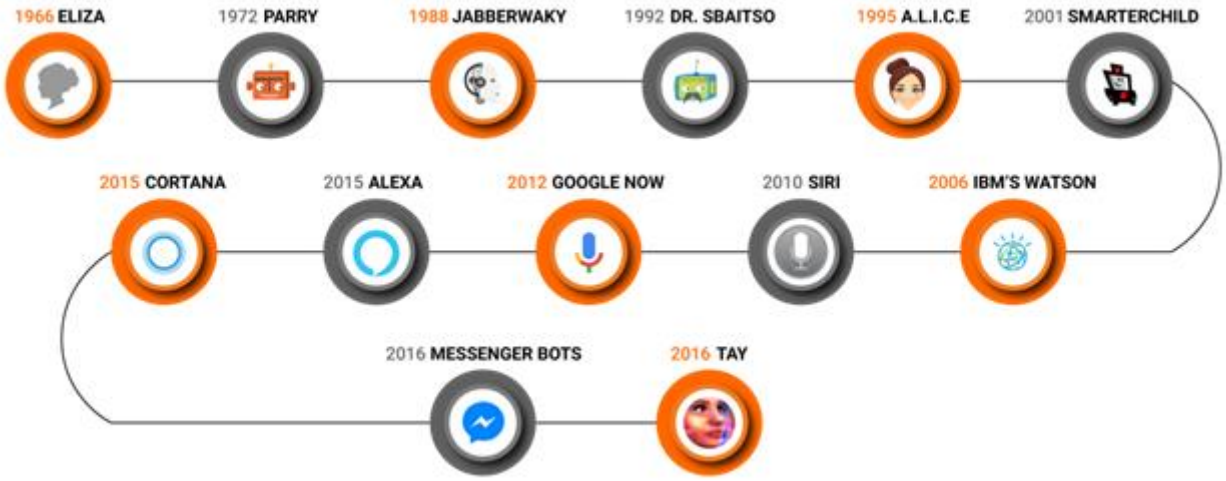


Figure 3. The History of Chatbot [8]

The early stages of chatbot development were marked by rule-based systems. These chatbots followed predefined scripts and responded to user inputs based on a set of predetermined rules. While these systems provided some level of automated interaction, their limited conversational abilities and lack of contextual understanding often resulted in frustrating user experiences [17].

The advent of Natural Language Processing (NLP) marked a significant turning point in the evolution of chatbots. NLP techniques enabled chatbots to understand and process human language, allowing for more natural and interactive conversations. Initially, keyword matching and pattern recognition were used to enhance chatbot responses. However, the limitations of these approaches became apparent when faced with complex queries and the need for deeper understanding.

With the rise of machine learning and deep learning, chatbots entered a new era of development. Advanced algorithms and neural networks began to play a crucial role in training chatbots to understand language nuances and generate more intelligent responses. This led to the emergence of chatbots powered by natural language understanding (NLU) models, which improved their ability to interpret user intent and extract relevant information from conversations.

One of the major breakthroughs in chatbot technology was the introduction of Generative Pre-trained Transformers (GPT) models. GPT, developed by OpenAI, revolutionized the capabilities of chatbots by leveraging large-scale language models. These models are pre-trained on vast amounts of text data and can generate human-like responses by capturing intricate language patterns and context. GPT-powered chatbots have significantly enhanced the conversational experience, providing more accurate and contextually relevant responses to user queries.

The evolution of chatbots has not only been driven by advancements in AI but also by the growing availability of data and computing power. The increased availability of large datasets and more powerful hardware has facilitated the training of more complex and accurate chatbot models. Additionally, the integration of chatbots with other technologies such as speech recognition, natural language generation, and machine vision has further expanded their capabilities, enabling them to handle a wider range of tasks and interact through multiple modalities.

Today, chatbots are employed in various industries and domains, including customer service, e-commerce, healthcare, and finance. They have become integral parts of businesses' customer support strategies, offering round-the-clock assistance, reducing response times, and enhancing customer satisfaction. Furthermore, chatbots have proved to be valuable tools for collecting user feedback, providing personalized recommendations, and automating routine tasks.

As chatbot technology continues to advance, future developments hold the promise of even more intelligent and context-aware virtual assistants. Ongoing research and innovation in AI, NLP, and machine learning will lead to improved language understanding, better contextual awareness, and enhanced personalization. Integration with emerging technologies like augmented reality and virtual reality may enable chatbots to provide immersive and interactive experiences.

III. CAPABILITIES AND LIMITATIONS OF GPT-POWERED CHATBOTS

GPT-powered chatbots, leveraging the power of Generative Pre-trained Transformers, have emerged as advanced virtual assistants capable of engaging in natural language conversations and delivering contextually relevant responses. These chatbots have showcased impressive capabilities that have transformed customer service interactions. However, it is important to acknowledge that they also come with certain limitations. This article explores the capabilities and limitations of GPT-powered chatbots, shedding light on their potential and areas for improvement.

One of the primary capabilities of GPT-powered chatbots lies in their ability to understand and generate natural language responses. These chatbots are pre-trained on large amounts of text data, enabling them to capture intricate language patterns and context. This allows them to engage in more meaningful and human-like conversations with users. GPT-powered chatbots excel in handling a wide range of customer inquiries, from basic questions to complex problem-solving. They can interpret user intent, extract relevant information, and provide accurate and timely responses, enhancing the customer experience and satisfaction.

Another key capability of GPT-powered chatbots is their potential for personalization. These chatbots can learn from user interactions and adapt their responses based on individual preferences and past conversations. By analyzing past interactions and understanding user preferences, GPT-powered chatbots can provide tailored recommendations, personalized offers, and a more customized experience to each user. This level of personalization strengthens customer engagement and fosters a sense of individualized attention.

GPT-powered chatbots also offer significant advantages in terms of efficiency and scalability. They can handle a large volume of inquiries simultaneously, reducing wait times and ensuring prompt responses to customer queries. With their 24/7 availability, GPT-powered chatbots enable round-the-clock customer support, eliminating time zone constraints and ensuring uninterrupted service. This scalability and efficiency relieve the burden on human agents, allowing them to focus on more complex and high-value tasks.

Despite their impressive capabilities, GPT-powered chatbots have certain limitations that need to be acknowledged. One notable limitation is their tendency to generate responses that may be irrelevant, misleading, or lack coherence. GPT models operate on statistical patterns learned from training data, which can sometimes lead to incorrect or nonsensical responses. Continuous training, fine-tuning, and human supervision are necessary to mitigate these issues and improve the quality of responses.

Furthermore, GPT-powered chatbots may struggle with understanding ambiguous queries or specialized domain knowledge. They rely heavily on the context provided in the conversation and may struggle to grasp the underlying intent in certain situations. Additionally, GPT models have limitations in their ability to handle complex multi-turn conversations seamlessly. Managing context and maintaining coherence across multiple user inputs can be a challenge for GPT-powered chatbots.

Ethical considerations also need to be addressed when deploying GPT-powered chatbots. These chatbots can inadvertently amplify biases present in the training data, leading to biased or discriminatory responses. Ensuring fairness and equity in chatbot interactions requires careful monitoring, bias detection, and proactive mitigation strategies [9].

IV. APPLICATION OF GPT-POWERED CHATBOTS IN INDUSTRIES

The application of GPT-powered chatbots has permeated various industries, revolutionizing the way businesses interact with their customers and improving overall operational efficiency. These advanced virtual assistants, driven by Generative Pre-trained Transformers models, have proven to be versatile and effective tools in providing personalized support, streamlining processes, and enhancing customer satisfaction. In this article, we explore the diverse applications of GPT-powered chatbots across industries.

Industry 4.0 is a name for the current trend of automation and data exchange in manufacturing technologies. It includes cyber-physical systems, the Internet of things, cloud computing and cognitive computing. Industry 4.0 creates what has been called a “smart factory”.

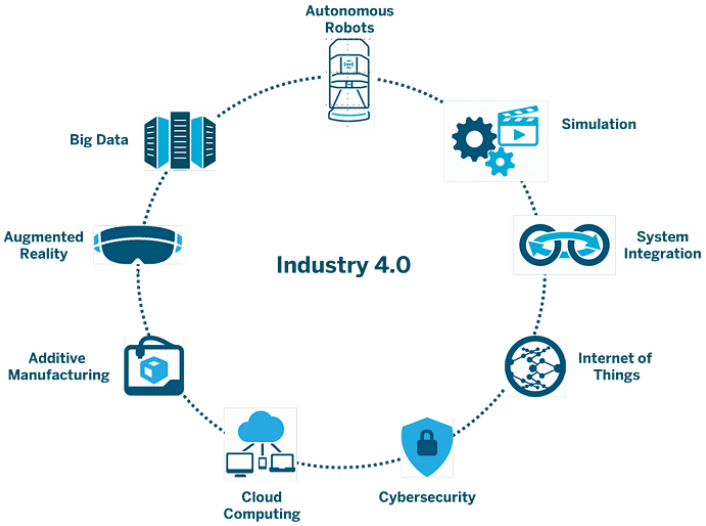


Figure 4. The Role of ChatBot in Industry 4.0 [10]

One industry where GPT-powered chatbots have made a significant impact is e-commerce. Chatbots integrated into e-commerce platforms can assist customers with product recommendations, answer inquiries about product details, and facilitate the purchasing process. By analyzing customer preferences and browsing history, chatbots can provide tailored suggestions, improving cross-selling and upselling opportunities. Additionally, they can handle order tracking, delivery status inquiries, and resolve customer issues, offering a seamless and convenient shopping experience.

The customer service sector has also witnessed a transformation with the application of GPT-powered chatbots. In this domain, chatbots can handle a wide range of customer queries, reducing the load on human agents and providing faster response times. They excel in tasks such as answering frequently asked questions, troubleshooting technical issues, and providing basic support. By automating routine tasks, chatbots allow human agents to focus on complex and specialized queries, enhancing overall efficiency and customer satisfaction.

The travel and hospitality industry has embraced GPT-powered chatbots to enhance customer experiences. Chatbots integrated into hotel websites or travel platforms can assist customers with room reservations, flight bookings, and travel recommendations. They can provide information on hotel amenities, local attractions, and restaurant suggestions. By offering personalized travel itineraries and real-time assistance, chatbots in this industry enhance customer engagement and contribute to a seamless and enjoyable travel experience.

Healthcare is another sector where GPT-powered chatbots have found valuable applications. These chatbots can provide initial triage and symptom assessments, offering guidance to patients before they consult a healthcare professional. They can answer queries about medications, provide reminders for medication schedules, and offer general health advice. GPT-powered chatbots in healthcare help alleviate the burden on healthcare providers and improve accessibility to basic medical information, particularly in remote or underserved areas.

In the banking and financial services industry, GPT-powered chatbots have transformed customer interactions. Chatbots integrated into banking apps or websites can assist customers with account inquiries, transaction history, and fund transfers. They can provide information on interest rates, loan options, and investment recommendations. By offering personalized financial insights and advice, these chatbots enhance customer engagement and simplify complex financial processes.

Education is another domain where GPT-powered chatbots have found applications. Chatbots can provide personalized tutoring, answer student queries, and offer guidance on assignments. They can also facilitate language learning by engaging in conversational practice with learners. GPT-powered chatbots in education contribute to individualized learning experiences and provide immediate feedback, enhancing the efficiency and effectiveness of learning processes [12].

V. EFFECTIVENESS OF GPT-POWERED CHATBOTS IN CUSTOMER SERVICE

GPT-powered chatbots, driven by the powerful Generative Pre-trained Transformers models, have emerged as highly effective tools in the realm of customer service. These AI-driven virtual assistants have revolutionized the way businesses interact with their customers, offering personalized support, improving response times, and enhancing overall customer satisfaction. In this article, we delve into the effectiveness of GPT-powered chatbots in customer service and examine how they have transformed the landscape of customer support.

One of the key areas where GPT-powered chatbots excel is their ability to handle a wide range of customer inquiries. These chatbots leverage their training on vast amounts of text data to understand natural language and interpret user intent accurately. Whether it's basic queries about product information, troubleshooting assistance, or complex issues, GPT-powered chatbots can provide relevant and timely responses. The speed and accuracy of their interactions contribute to a more efficient and satisfactory customer service experience.

Furthermore, GPT-powered chatbots offer personalized support to customers. Through their analysis of previous interactions and user data, these chatbots can tailor their responses and recommendations based on individual preferences and past conversations. This personalization enhances the customer experience by making interactions more relevant and engaging. Customers feel understood and valued when chatbots can address their specific needs, leading to increased satisfaction and brand loyalty.

The round-the-clock availability of GPT-powered chatbots is another significant factor contributing to their effectiveness in customer service. Unlike human agents, chatbots can operate 24/7, providing support and assistance at any time of the day. This availability ensures that customers receive immediate responses to their queries, regardless of their time zone or location. The elimination of wait times and the ability to resolve issues promptly contribute to a positive customer experience and a sense of reliability and convenience.

GPT-powered chatbots also play a pivotal role in improving the efficiency of customer service operations. By automating routine and repetitive tasks, these chatbots free up human agents to focus on more complex and value-added activities. Chatbots can handle a large volume of inquiries simultaneously, ensuring that customers are served promptly. This efficiency translates into cost savings for businesses, as they can optimize their resources and reduce the need for additional staffing in customer service departments.

Moreover, GPT-powered chatbots contribute to knowledge management and self-service support. They can access vast repositories of information, including product manuals, FAQs, and troubleshooting guides, and provide customers with accurate and up-to-date information. By empowering customers to find answers and solutions on their own, chatbots reduce the need for human intervention and enable self-service support. This empowers customers, saves time for both customers and agents.

VI. ETHICAL CONSIDERATIONS AND USER TRUST

The rise of GPT-powered chatbots in various industries has brought about numerous benefits, but it also raises important ethical considerations and challenges related to user trust. As these chatbots leverage powerful language models to generate responses, it becomes crucial to address potential issues such as bias, privacy, transparency, and the preservation of user trust. In this article, we delve into the ethical considerations surrounding GPT-powered chatbots and the importance of fostering user trust.

One of the foremost ethical concerns in deploying GPT-powered chatbots is the potential for bias. These language models are trained on vast amounts of data from the internet, which may contain inherent biases. As a result, chatbot responses can inadvertently reflect and perpetuate these biases. It is essential for developers to actively monitor and address bias in training data, ensuring fairness and inclusivity in chatbot interactions. By actively working to reduce biases and promote diversity, chatbot developers can foster a more equitable user experience.

Another significant ethical consideration is privacy. GPT-powered chatbots often process and store user interactions and personal data to improve their performance. It is crucial to implement robust data protection measures and adhere to data privacy regulations to safeguard user information. Transparency regarding data collection, storage, and usage is essential in establishing trust with users. Clearly communicating privacy policies and obtaining user consent for data processing helps to maintain user trust and respect their privacy.

Transparency in chatbot interactions is another critical aspect of ethical considerations. Users should be informed when they are interacting with a chatbot and understand the limitations of its capabilities. Clearly

distinguishing between human agents and chatbots helps to manage user expectations and ensures transparency in the communication process. Additionally, it is important to be transparent about the chatbot's limitations and clearly communicate its capabilities to avoid user frustration or misinformation.

Ensuring user trust in GPT-powered chatbots is vital for their successful adoption. Users need to feel confident that their interactions are secure, confidential, and reliable. Designing chatbot interfaces that provide clear and accurate information, maintaining consistency in responses, and ensuring error handling protocols are essential for building user trust. Regularly updating and improving chatbot performance through user feedback and ongoing development efforts also contribute to enhancing user trust [19].

Furthermore, it is important to consider the potential impact of GPT-powered chatbots on employment and human labor. While chatbots can automate routine tasks, it is crucial to balance automation with the preservation of human jobs and dignity. Companies should have transparent policies regarding the deployment of chatbots and actively engage with employees to ensure a smooth transition and provide opportunities for upskilling and reassignment.

VII. CHALLENGES AND LIMITATIONS

GPT-powered chatbots, driven by advanced natural language processing models, have demonstrated remarkable capabilities in engaging with users and providing personalized responses. However, like any technology, they also come with certain challenges and limitations that need to be addressed. In this article, we explore the challenges and limitations of GPT-powered chatbots, shedding light on areas that require further development and improvement.

One of the key challenges of GPT-powered chatbots lies in their potential to generate irrelevant or nonsensical responses. These chatbots operate based on statistical patterns learned from training data, which means that they might generate answers that seem plausible but are not contextually appropriate or accurate. Fine-tuning and continuous training are necessary to refine the chatbot's responses and improve its ability to generate coherent and relevant replies. The challenge lies in striking the right balance between generating creative responses and maintaining accuracy and coherence.

Contextual understanding is another significant challenge for GPT-powered chatbots. While these chatbots can handle single-turn conversations effectively, they can struggle with managing complex multi-turn interactions. Maintaining context across multiple user inputs and understanding the dependencies between various messages can be challenging. Enhancing the chatbot's ability to track and remember the conversation history accurately is crucial to ensuring a seamless and coherent user experience.

The limitation of GPT-powered chatbots in handling ambiguous queries is also noteworthy. Ambiguity in user queries can arise due to misspellings, typos, or vague language. GPT models rely heavily on the context provided, and when faced with ambiguous queries, they may struggle to understand the intended meaning. Improving the chatbot's ability to ask clarifying questions or provide suggestions when faced with ambiguity is essential for enhancing user satisfaction and reducing frustration.

Another limitation of GPT-powered chatbots is their lack of specialized domain knowledge. While these chatbots can generate responses based on their training data, they may struggle when faced with queries or topics outside their training scope. Incorporating domain-specific knowledge and integrating external databases

or APIs can help overcome this limitation and enhance the chatbot's ability to provide accurate and comprehensive information in specialized domains.

Furthermore, ethical considerations and bias pose significant challenges for GPT-powered chatbots. As these chatbots learn from vast amounts of text data, they can inadvertently reflect and perpetuate biases present in the training data. Ensuring fairness, inclusivity, and mitigating biases in chatbot responses require ongoing monitoring, bias detection, and continuous improvement. Developers need to actively address these ethical concerns and promote transparency and fairness in chatbot interactions.

Lastly, user trust is a critical challenge for GPT-powered chatbots. Building and maintaining user trust is crucial for the successful adoption and acceptance of chatbot technology. Users need to feel confident that their interactions are secure, private, and reliable. Addressing privacy concerns, providing clear communication about the chatbot's capabilities and limitations, and ensuring consistent and accurate responses are essential for fostering user trust in GPT-powered chatbots.

VIII. FUTURE DIRECTIONS AND OPPORTUNITIES

GPT-powered chatbots have made significant strides in enhancing customer experiences, automating tasks, and improving operational efficiency. Looking ahead, there are several exciting future directions and opportunities for the development and application of these advanced virtual assistants. In this article, we explore some of the potential avenues for the future of GPT-powered chatbots.

One promising area for future development is the advancement of conversational abilities. While GPT-powered chatbots have shown impressive language generation capabilities, there is room for improvement in terms of creating more engaging and interactive conversations. Enhancing the chatbot's ability to understand nuanced user intents, handle complex multi-turn conversations, and generate more dynamic and contextually appropriate responses will be crucial. This can involve the incorporation of reinforcement learning techniques and the development of more advanced dialogue management systems.

Another exciting direction for GPT-powered chatbots is their integration with emerging technologies. Chatbots that can seamlessly integrate with voice assistants, augmented reality (AR), or virtual reality (VR) platforms can enhance user experiences and offer more immersive interactions. For example, users could have natural and interactive conversations with chatbots through voice commands, or receive visual information overlay through AR/VR interfaces. The combination of GPT-powered chatbots with these technologies opens up new possibilities for innovative applications in various industries [18].

Improving the chatbot's ability to handle real-time data and dynamic environments is another area of opportunity. GPT-powered chatbots have primarily been trained on static text data, limiting their ability to provide up-to-date and contextually relevant information in rapidly changing scenarios. Integrating real-time data sources, such as news feeds or social media streams, can enable chatbots to stay informed and provide timely and accurate responses. This can be particularly valuable in industries where staying updated with the latest information is crucial, such as finance, news, or emergency services.

The future of GPT-powered chatbots also lies in their enhanced personalization capabilities. By leveraging user data, preferences, and historical interactions, chatbots can offer highly personalized and tailored experiences. This can involve recommending products or services based on individual preferences, adapting to user communication styles, or even incorporating emotional intelligence to understand and respond empathetically

to user emotions. Personalized chatbot interactions can significantly enhance customer satisfaction and engagement, leading to stronger customer relationships [[19,20,21].

Furthermore, the ethical considerations surrounding GPT-powered chatbots will continue to be a focus in the future. Developers will strive to address biases, ensure data privacy and security, and promote fairness and transparency in chatbot interactions. Research and development efforts will be dedicated to improving bias detection and mitigation techniques, implementing stricter privacy measures, and fostering user trust through clear communication and accountability.

IX. CASE STUDIES: SUCCESSFUL IMPLEMENTATIONS OF GPT-POWERED CHATBOTS

GPT-powered chatbots have revolutionized the way we interact with technology, and two prime examples of successful implementations are Amazon Alexa and Google Assistant. These virtual assistants have become household names, demonstrating the immense potential and impact of GPT-powered chatbots in our daily lives. In this article, we delve into the successful implementations of Amazon Alexa and Google Assistant, highlighting their key features and the benefits they bring to users.

Amazon Alexa, powered by GPT technology, has transformed the way we interact with our homes. As a voice-activated virtual assistant, Alexa seamlessly integrates with various smart devices, enabling users to control their smart home appliances, play music, set reminders, and access a wide range of information and services using simple voice commands. The success of Alexa lies in its ability to understand and respond accurately to natural language queries, making it an intuitive and user-friendly assistant. With a vast library of skills and continuous updates, Alexa continues to evolve and offer new functionalities, enhancing the user experience and making daily tasks more convenient and efficient [13].

Similarly, Google Assistant, powered by GPT technology, has made significant strides in providing personalized and contextual assistance. Google Assistant is available on a variety of devices, including smartphones, smart speakers, and smart displays. It offers users a seamless and conversational experience, allowing them to perform tasks, ask questions, receive recommendations, and get real-time information. Google Assistant leverages its deep integration with Google's vast knowledge graph, search engine, and services to provide accurate and comprehensive responses. The assistant's ability to understand user preferences and adapt to individual needs contributes to a highly personalized experience, fostering stronger user engagement and satisfaction.

The successful implementations of Amazon Alexa and Google Assistant highlight the benefits that GPT-powered chatbots bring to users. These virtual assistants offer convenience, efficiency, and a hands-free approach to accessing information and performing tasks. They simplify daily routines, provide entertainment, and act as a central hub for controlling smart home devices. The ability to have natural language conversations with these chatbots enhances the user experience, making technology more accessible and intuitive [11].

Moreover, GPT-powered chatbots like Alexa and Google Assistant continuously evolve through regular updates and integrations with third-party services. This allows them to expand their capabilities and offer a wider range of functionalities, ensuring they remain relevant and valuable to users. The success of these implementations has also fostered a vibrant developer community, with individuals and companies creating new skills and actions that further enhance the capabilities of these virtual assistants.

X. CONCLUSION

In conclusion, the impact of chatbots powered by GPT on customer service is undeniably significant. The capabilities of GPT-powered chatbots have transformed the customer service landscape, providing businesses with efficient and effective tools to enhance customer experiences, streamline operations, and improve overall satisfaction. Through a comprehensive review of the topic, several key insights and trends have emerged.

Firstly, the evolution of chatbots, driven by advancements in artificial intelligence, has been remarkable. GPT-powered chatbots have come a long way in terms of their conversational abilities, natural language processing capabilities, and context-awareness. They have evolved from simple rule-based systems to sophisticated virtual assistants capable of understanding and generating human-like responses.

Secondly, the effectiveness of GPT-powered chatbots in customer service has been demonstrated through various case studies and real-world implementations. From e-commerce customer support to healthcare triage and financial services, these chatbots have proven their value in automating tasks, providing personalized assistance, and improving response times. They have contributed to higher customer satisfaction, increased efficiency, and reduced workload on human agents.

Furthermore, the application of GPT-powered chatbots extends beyond customer service. They have found utility in industries such as finance, travel, and hospitality, where they assist with tasks ranging from financial advice to hotel reservations. The versatility and adaptability of these chatbots make them valuable assets across diverse sectors.

While the benefits of GPT-powered chatbots are evident, ethical considerations and user trust remain important focal points. Developers and researchers are actively addressing biases, privacy concerns, and fairness in chatbot interactions. They are working towards improving bias detection and mitigation techniques, implementing stricter privacy measures, and promoting transparency to foster user trust.

Looking to the future, there are exciting opportunities and challenges for GPT-powered chatbots. Advancements in conversational abilities, integration with emerging technologies, real-time data handling, and personalized interactions hold immense potential. By leveraging these opportunities and adhering to responsible development practices, GPT-powered chatbots will continue to revolutionize industries, enhance customer experiences, and shape the future of human-computer interactions.

In conclusion, the impact of chatbots powered by GPT on customer service is profound. They have reshaped the way businesses engage with customers, offering personalized assistance, improving efficiency, and enhancing overall satisfaction. With ongoing advancements and a commitment to ethical considerations, GPT-powered chatbots will undoubtedly play a crucial role in the future of customer service, driving innovation, and delivering exceptional experiences.

XI. REFERENCES

[1]. S.S. Biswas, Potential use of chat GPT in global warming, Ann. Biomed. Eng. (2023) 1–2.

[2]. S.S. Biswas, Role of chat GPT in public health, Ann. Biomed. Eng. (2023) 1–2.

[3]. R.W. McGee, Annie Chan: Three Short Stories Written with Chat GPT, 2023. Available at: SSRN 4359403.

[4]. R.W. McGee, Is chat GPT biased against conservatives? An empirical study, Empir.Stud. (2023) (February 15, 2023).
[5]. A. Mathew, Is artificial intelligence a world changer? A case study of OpenAI's chat GPT, Recent Prog. Sci. Technol. 5 (2023) 35–42
[6]. https://itchronicles.com/artificial-intelligence/the-rise-of-chatbots/
[7]. How to Choose the Best Chatbot for Your E-Commerce Business (livechat.com)
[8]. How AI Chatbot Is Transforming Customer Care industry? (automationedge.com)
[9]. Sakib Shahriar,Kadhim Hayawi,Let's have a chat! A Conversation with ChatGPT: Technology, Applications, and Limitations(2023)
[10]. The Role of ChatBot in Industry 4.0 by Aydın Fevzi Özçekiç & Chatbots Life
[11]. W. Jiao, W. Wang, J. Huang, X. Wang, and Z. Tu, "Is ChatGPT A Good Translator? A Preliminary Study." arXiv, Jan. 31, 2023. doi: 10.48550/arXiv.2301.08745
[12]. D. Mhlanga, "Open AI in Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning." Rochester, NY, Feb. 11, 2023. doi: 10.2139/ssrn.4354422.
[13]. B. D. Lund and T. Wang, "Chatting about ChatGPT: how may AI and GPT impact academia and libraries?," Libr. Hi Tech News, vol. ahead-of-print, no. ahead-of-print, Jan. 2023, doi: 10.1108/LHTN-01-2023-0009.
[14]. E. Kasneci et al., "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education." EdArXiv, Jan. 29, 2023. doi: 10.35542/osf.io/5er8f.
[15]. H. H. Thorp, "ChatGPT is fun, but not an author," Science, vol. 379, no. 6630, pp. 313–313, Jan. 2023, doi: 10.1126/science.adg7879.
[16]. A. Radford et al., "Language models are unsupervised multitask learners," OpenAI Blog, vol. 1, no. 8, p. 9, 2019.
[17]. A. Følstad and P. B. Brandtzaeg, "Users' experiences with chatbots: findings from a questionnaire study," Qual. User Exp., vol. 5, no. 1, p. 3, Apr. 2020, doi: 10.1007/s41233-020-00033-2.
[18]. E. Adamopoulou and L. Moussiades, "An Overview of Chatbot Technology," in Artificial Intelligence Applications and Innovations, Cham, 2020, pp. 373–383. doi: 10.1007/978-3-030- 49186-4_31.
[19]. B. Luo, R. Y. K. Lau, C. Li, and Y.-W. Si, "A critical review of state-of-the-art chatbot designs and applications," WIREs Data Min. Knowl. Discov., vol. 12, no. 1, p. e1434, 2022, doi: 10.1002/widm.1434.
[20]. T.-J. Chen, "ChatGPT and other artificial intelligence applications speed up scientific writing," J. Chin. Med. Assoc., p. 10.1097/JCMA.0000000000000900
[21]. S. Jalil, S. Rafi, T. D. LaToza, K. Moran, and W. Lam, "ChatGPT and Software Testing Education: Promises & Perils." arXiv, Feb. 08, 2023. doi: 10.48550/arXiv.2302.03287.
[22]. B. Guo et al., "How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection." arXiv, Jan. 18, 2023. doi: 10.48550/arXiv.2301.07597.

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Applications, Methods, And Trends in A Comprehensive Review of IoT in Healthcare

Wasim Yasin[1], Pavan Kumar R B[1]
[1]Assistant Professor
Department of Computer Science, East Point College of Engineering and Technology, Bengaluru, Karnataka, India

ABSTRACT
The Internet of Things (IoT) has shown great promise in revolutionizing various industries, including healthcare. The Healthcare Internet of Things (HIoT) integrates software, hardware, physical objects, and computing devices to enable seamless communication, data collection, and exchange between humans and various physical and virtual healthcare resources. This paper presents a systematic review of 146 articles published between 2015 and 2020 that investigate the applications of HIoT in healthcare systems. The review aims to identify, compare, and classify the existing investigations taxonomically, leading to the development of a comprehensive taxonomy for HIoT. The articles are analyzed technically and classified into five categories, namely sensor-based, resource-based, communication-based, application-based, and security-based approaches. The paper also discusses the benefits and limitations of the selected methods, providing a comprehensive comparison in terms of evaluation techniques, evaluation tools, and evaluation metrics.
Keywords: Internet of things (IoT) Healthcare e-health Systematic review.

I. INTRODUCTION

One of the primary advantages of incorporating IoT in remote health management is the ability to collect real-time health data from patients through various wearable devices and sensors. These devices can monitor vital signs, physical activity, sleep patterns, and other health-related parameters, providing healthcare professionals with valuable insights into a patient's condition without the need for in-person visits. This remote monitoring capability is particularly beneficial for elderly and disabled individuals who may have difficulty traveling to healthcare facilities frequently.

Moreover, IoT-enabled remote health management systems allow for personalized and targeted healthcare interventions. By continuously collecting and analyzing patient data, healthcare providers can tailor treatment plans and interventions based on individual needs and specific health conditions. This personalized approach can lead to better health outcomes and improved patient satisfaction.

Another significant advantage of IoT in healthcare is its potential to reduce healthcare costs. By enabling remote monitoring and early detection of health issues, IoT can help prevent hospitalizations and emergency

room visits, which are often costly. Additionally, remote health management can lead to more efficient use of healthcare resources and reduce the burden on healthcare facilities.

However, the widespread adoption of IoT in healthcare also poses some challenges and concerns. One major concern is data privacy and security. As sensitive health data is transmitted and stored across various devices and networks, there is a risk of unauthorized access and data breaches. Healthcare organizations and technology providers must implement robust security measures to safeguard patient information and ensure compliance with data protection regulations.

Furthermore, the integration of IoT devices and technologies into existing healthcare systems may require substantial investments in infrastructure, training, and maintenance. Ensuring seamless interoperability among different IoT devices and platforms is crucial for the success of remote health management systems.

II. LITERATURE SURVEY

Numerous scholars have contributed research papers and tackled numerous chronic disease challenges, including AI in healthcare and IoT in healthcare. A section emphasizes a review of the relevant literature.

Shima Okada, NaruhiroShioza et. al[7] focused their attention mostly on bodily motions while they slept since they are most closely associated to the sleep-wake cycle. They have put out a method that makes use of several image processing techniques to measure body motions while you sleep. They used video monitoring to characterize the differences in body movement during sleep in normal children and those with ADHD in order to prove the validity of the proposed system. Body movement data calculated by difference image processing were compared with the sleep stages measured by PSG.

A brand-new brain-computer interface (BCI) device designed to treat children with attention deficit hyperactivity disorder was introduced by Darius Adam Rohani, Helge B.D, and colleagues [8]. It uses the P300 potential in a series of feedback games to improve the subjects' attention. To identify the disorder, they used a support vector machine (SVM) that applied temporal and template-based features.

Tao Liu and Dongxin Lu [10] provide a conceptual overview of IoT (the internet of things), as well as information on its history, various interpretations, major technology, and applications.

Dong-Hwan Park and Hyo-Chan Bang[11] concentrated on how technologies help to improve the interoperability of IoT devices and make usage of IoT devices more conveniently. IoT device interoperability and semantic-based information services are made possible by the expected platform technology. This service platform can be used for many semantic IoT services, such as the distribution, participation, and sharing of open sensing data, the collection of invisible information in tangible environments by smart devices, and the provision of smart living services.

A remote health care monitoring system was proposed by Sai Kiran P, Rajalakshmi, and others [13], with the assembled medical data from biomedical sensors being transmitted to the adjacent gateway for auxiliary processing. Data transmission contributes significantly to transmitter power consumption and an increase in network traffic. The amount of electricity saved and the decline in network traffic are the measures left over for performance study. The suggested rule engine significantly lowers the amount of energy used and the amount of network traffic produced.

A semantic data paradigm for storing and interpreting IoT data was proposed by Boyi Xu, Li Da Xu, and others [15]. Then, in order to increase the availability of IoT data resources, a resource-based ubiquitous data accessing

technique is created and designed. It allows for the universal acquisition and use of IoT data. Last but not least, they presented an Internet of Things (IoT)-based system for emergency medical services and showed how to gather, integrate, and interoperate IoT data.

Nuno Vasco Lopes, Filipe Pintoz, and colleagues [16] presented an IoT architecture for disabled individuals that aims to describe, select the additional most appropriate IoT technologies, and include global standards into the design of the suggested architecture. They focus on the empowering IoT technologies and their viability for those with disabilities.

III. PURPOSE OF IOT IN HEALTHCARE

If necessary, preprocessing of human data collected. Ease of access to Internet, wearable IoT devices have beena huge hit in the market. Because end users, clients andcustomers in healthcare network are humans (patients orhealth-conscious individuals), developing of ambient intelligence is crucial.To provide Ambient intelligence for the continuous learning about patient's data executes any required actiontriggered by a recognized event. The Integration ofautonomous control and human computer interaction (HCI)technologies into ambient intelligence can further enhancethe capability of IoT-aided healthcare services.To gain the information about human in real time through IoT wearable device.Preprocessing of data acquisitioned about human (ifnecessary). Using data mining tools to analyze and predict chronic diseases at an early stage, which offers decision-making approach. To provide healthcare solutions based on IoT wherever, anytime.

IV. PROPOSED METHODOLOGY

The information and knowledge for the prognosis of chronic illnesses can be inferred from wearable health care equipment. Tier-1 unprocessed and raw data is obtained from a wearable IoT healthcare monitoring device that has a variety of sensors, including an EEG sensor, a galvanic skin response sensor, an ECG sensor, an accelerometer, and a skin temperature sensor, among others. If there is a large amount of existing irrelevant and redundant information or noisy and unreliable data, information is inferred from the data at layer 2 by filtering, processing, categorizing, condensing, and contextualizing the data. To predict chronic diseases (such as cardiovascular diseases, mental health disorders, diabetes, stroke, etc.) in tier three, we must develop algorithms using a variety of mining techniques, including constraint-based mining and periodic mining.

V. THE STATE OF IOT HEALTHCARE AND SEVERAL WELL-KNOWN TECHNOLOGY COMPANIES' FUTURE PLANS

The extent of IoT use in business and government is described in this section:

Windows and Google: In order to facilitate code sharing, Google, Microsoft has concentrated on employing an intelligent system to uncover the promise of IoT-based health care solutions. Source physical web standard for IoT may be considered an attempt to arrange an easier technique to communicate with connected medical equipment [17].

Intelligent systems serve as the skeleton for the technologies that enable the collection of health data from gadgets while maintaining the necessary connectivity [18].

IBM and Intel For many existing medical workflow environments, Intel places a strong emphasis on real-time synchronous communication systems and health data streaming, which can assist shorten cycle times and enhance first-time quality.

Through the idea of "smarter health care," IBM redefines value and success in the healthcare industry. IBM has collaborated with other well-known companies throughout the world to produce a collection of IoT devices. It focuses on a number of healthcare solutions, including connected home health, data governance for the healthcare industry, and healthcare providers' access to health analytics[20].

Apple: Apple has declared the Internet of Things to be the most advanced technology. The apple watch is a fitness tracker, smart watch, and heart rate monitor. For effective and connected healthcare services concentrating on secure access, physician gains, and improved care, the Memorial Hermann hospital system is wholly dependent on Apple's Solutions [21].

Qualcomm and Cisco: Cisco is prepared to offer convergent solutions based on unrelated technologies.

Advanced data analytics, networks, and can introduce efficient algorithms for handling cumulative traffic loads arising from widely used IoT healthcare devices. The 2net Platform of Qualcomm Life offers a set of wireless health solutions that can capture and deliver health device data from almost all users' wireless medical devices to integrated portals and databases. CISCO is collaborating with top healthcare organizations to develop medical-grade network architecture [22]. IoT healthcare solutions must be advanced and innovative, according to Qualcomm [23]. the public sector Additionally, the Indian government launched numerous programs to promote IoT in healthcare. These initiatives to increase IoT use in India's healthcare industry are to be expected. The majority of the nation favors the United States, Japan, Germany, Korea, China, France, and Australia.

VI. CONCLUSION

Any problem or medical ailment that is early identified can be most effectively treated. It is evident from the literature mentioned above that the main objective of the research must be The collection, management, and application of information in health data via IoT has significantly improved the quality and efficiency of healthcare as well as the ability to react to widespread public health situations.

VII. REFERENCES

[1]. D.T.Felson, A.Naimark, J.Anderson, L.Kazis, W.Castelli, and R.F Meenan, ``The prevalence of knee osteoarthritis in the elderly. The framingham osteoarthritis study,'' Arthritis Rheumatism, vol. 30, no. 8, pp. 914_918, Aug. 1987, doi: 10.1002/art.1780300811.

[2]. T.Vos et al., ``Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013,'' Lancet, vol. 386, no. 9995, pp. 743_800, 2015,doi: 10.1016/S0140-6736(15)60692-4.

[3]. L.Murphy and C.G.Helmick, ``The impact of osteoarthritis in the United States: A population-health perspective,'' Amer. J. Nursing, vol. 112, no. 3,pp. S13-S19, 2012.

[4]. S.Kurtz, K.Ong, E.Lau, F.Mowat, and M.Halpern, ``Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030,'' J. Bone Joint Surg., vol. 89, no. 4, pp. 780-785, Apr. 2007,doi: 10.2106/JBJS.F.00222.

[5]. K.R.Kaufman, C.Hughes, B.F.Morrey, M.Morrey, and K.-N. An, ``Gait characteristics of patients with knee osteoarthritis,'' J.Biomech., vol. 34, no. 7, pp. 907-915, 2001, doi:10.1016/s0021-9290(01)00036-7.

[6]. W.Yang, K.Wang, and W.Zuo, ``Neighborhood component feature selection for high-dimensional data,'' J.Compute., vol. 7, no. 1, pp. 161-168, Jan.2012

[7]. Steven J Soalding, C Kent Kwoh, Robert Boudreau, Joseph Enama, Julie Lunich, Daniel Huber, Louis Denes and RaphaelHirsch,3D and thermal surface imaging produces reliable measures of joint shape and temperature: a potential tool for quantifying arthritis, Research & Therapy, 2008 10: R10.

[8]. M. Hall, K. L. Bennell, T.V.Wrigley, B.R.Metcalf, P. K. Campbell, J. Kasza, K. L. Paterson, D. J. Hunter, and R.S.Hinman, ``The knee adduction moment and knee osteoarthritis symptoms: Relationships according to radiographic disease severity,'' Osteoarthritis Cartilage, vol. 25, no. 1,pp. 34-41, Jan. 2017, doi: 10.1016/j.joca.2016.08.014.

[9]. Beyer T, Freudenberg LS, Czernin J, and Townsend DW, —The future of hybrid imaging – part 3: PET/MR, small animal imaging and beyond, ‖ Insights Imaging, 2011, 2:235-246.

[10]. Chen Lin, Elizabeth W.Karlson, Helena Canhao,—Automatic Prediction of Rheumatoid Arthritis Disease Activity from the Electronic Medical Records, PLoS ONE 8(8): e69932. doi:10.1371/ journal. pone.

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Analysis of Artificial Intelligence Developments and Their Impacts on Cybersecurity

Shammi L[1], Dr. R Senkamalavalli[2], Manimegalai A[1]
[1]Assistant Professor, [2]Associate Professor
Department of Computer Science and Engineering, East Point College of Engineering and Technology, Karnataka, India

ABSTRACT
Artificial Intelligence (AI) is a potent tool that cybersecurity teams may use to automate repetitive processes, speed up threat detection and response, and increase the efficacy of their actions to fortify the security posture against a variety of security problems and cyberattacks. According to experts, Machine Learning (ML) and artificial intelligence both have beneficial and harmful consequences on cybersecurity. AI algorithms learn how to react to various scenarios using training data. They pick up new information by replicating and supplementing it along the way. This article examines AI's effects on cybersecurity, which are beneficial as well as detrimental.

Keywords: Cybersecurity, AI, Cyber threats, Artificial Intelligence, Machine Learning

I. INTRODUCTION

Networks, devices, Programmes, and data are protected by networks, processes, and practices known as cybersecurity from intrusions, damage, and unauthorized access. "Cybersecurity refers to the set of activities and measures, both technical and non-technical, intended to protect the 'real geography' of cyberspace as well as devices, software, and the information they contain communicated, from all possible threats," according to the definition given by Myriam Dunn Cavelty [3]. One of the most crucial challenges in cyberspace nowadays is cybersecurity [4, 5].

The 20th century saw the development of artificial intelligence. In an effort to design a structure that wouldn't need a human brain's assistance, this development came about. More research was done on the subject as a result of the discovery. Robot and intelligent system development have increased. The advances all made an effort to integrate a piece of technology that acts like a person but mimics human behaviour. Numerous mathematicians attempted to create formulas to aid with the element of the research, which also included mathematics. AI platforms assist enterprises in the development, management, and deployment of machine learning and deep learning models at scale. Decreasing software development tasks such as data management and deployment make AI technology more accessible and economical [2]. With the increase in cyber risks, artificial intelligence (AI) is increasingly widely employed to monitor and restrict cybercrime.

II. OVERVIEW OF ARTIFICIAL INTELLIGENCE

AI platforms let businesses create, manage, and deploy machine learning and deep learning models at scale. As businesses compete to incorporate AI technology into their goods, business strategies, or security programmes, the impact of AI on cybersecurity is swiftly growing in importance. The topic of artificial intelligence is increasingly becoming one that has the potential to revolutionize cybersecurity. However, the use of AI to cybersecurity poses fresh problems and hazards in addition to offering fresh and creative solutions. In order to employ AI in cybersecurity responsibly and ethically, it is crucial to grasp both its potential and its limitations.
AI technologies can understand, learn, and act based on the information derived from events and effects. According to Stuart Russell and Peter Norvig, "AI attempts not just to understand but also to build intelligent entities" and they offered a definition for AI, organized into two main categories, such as:

- thought process and reasoning: these measure success in terms of thinking, which is categorized into thinking humanly and thinking rationally.
- behavior: this measures a success based on the ideal performance and action, and it is categorized into acting humanly and acting rationally.

AI works in three ways [2]:

- Assisted intelligence, which improves what people are already doing
- Augmented intelligence, which empowers people to do things that they could not do
- Autonomous intelligence, which are features of machines that act on their own.

With respect to these three categories, it could be concluded that AI aims to solve some of the most difficult problems and cybersecurity falls into this category, since cyberattacks have become highly sophisticated and potentially more disastrous and turned to be a complex issue in cyberspace.

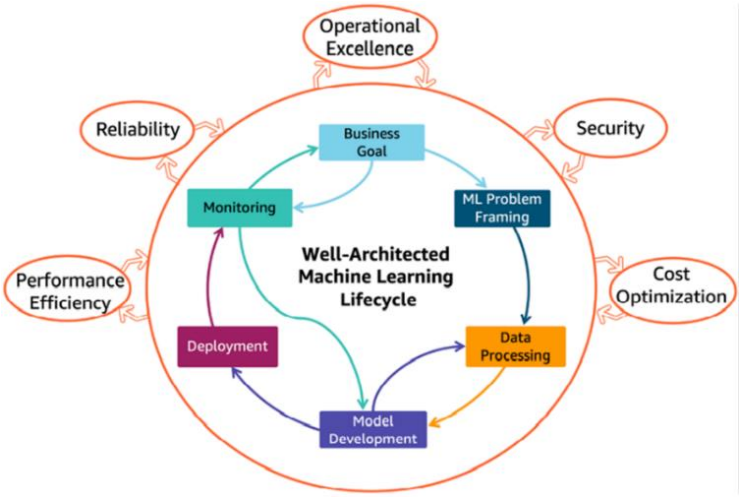


Fig:1 Machine Learning Life Cycle [4]

The greatest serious threat to corporations, organisations, and governments around the world is cyberattacks, which are continually being fought against by cyber security and intelligence firms globally. Cyberattacks are growing more frequent and effective as technology advances and makes life easier and more convenient every day. The huge arsenals that cybercriminals and malevolent third parties possess enable them to launch large-scale attacks.

III. AI CONTRIBUTION IN CYBERSECURITY

Artificial Intelligence has many applications in different sectors and industries. One of the sectors that have continued to benefit from artificial intelligence has been cybersecurity. In order to better understand the notion of various applications of AI for cybersecurity, this section analyses the background data pertaining to the core topics of this review, including the operational definition of cybersecurity and the AI taxonomy.

One of the many applications for artificial intelligence is cybersecurity. According to a Norton research, the average data breach recovery costs $3.86 million globally. According to the survey, it takes businesses 196 days on average to recover from a data breach. Organisations should increase their AI spending in order to prevent time wastage and financial losses. In data security measures and protocols, AI technology has been one of the best technologies in ensuring that they have improved them. Data is critical to business organizations, so it needs to be secured [6]. With the help of multiple data encryption protocols, the system can facilitate great encryption and guarantee the security of the data involved. The great protocol significantly impacts the technology in the cybersecurity sector of technology [2].

Threat intelligence, AI, and machine learning can identify trends in data to help security systems learn from the past. Additionally, AI and machine learning help businesses adhere to security best practises and speed up incident reaction times. In order to safeguard information and communication systems and the data they contain from harm, unauthorised use or modification, or exploitation, cybersecurity policies, procedures, and technical measures are put in place. The problem is made more difficult by the quickening rate of technological advancement and innovation as well as the rapidly changing nature of cyber threats. AI-based cybersecurity tools have evolved to assist security teams in effectively reducing risks and enhancing security in response to this unprecedented challenge. A generally acknowledged and streamlined taxonomy is required to review the research on using AI for cybersecurity due to the heterogeneity of AI and cybersecurity. Threats are recognised using signatures or indicators of compromise in traditional security procedures. This method may be successful against threats that have already been experienced, but it is ineffective against threats that have not yet been identified.

About 90% of threats [1] may be detected using signature-based strategies. Artificial intelligence (AI) can improve detection rates up to 95%, however there will be a huge number of false positives. The best course of action would be to use both conventional techniques and AI. By doing this, false positives can be reduced and the detection rate can reach 100%. Businesses can also utilise behavioural analysis in conjunction with AI to improve the threat hunting process. By processing large amounts of endpoint data, for instance, you can use AI models to create profiles of each application within a network for an organisation.

Fraud Prevention Techniques Using Artificial Intelligence: Fraud attempts and breaches are more nuanced, with organized crime and state-sponsored groups using machine learning algorithms to find new ways to defraud digital businesses.[9]. Fraud-based attacks have a completely different pattern, sequence, and structure, which make them undetectable using rules-based logic and predictive models alone.

The future of AI-based fraud prevention relies on the combination of supervised and unsupervised machine learning. Supervised machine learning excels at examining events, factors, and trends from the past. Historical data trains supervised machine learning models to find patterns not discernible with rules or predictive analytics. Unsupervised machine learning is adept at finding anomalies, interrelationships, and valid links

between emerging factors and variables. Combining both unsupervised and supervised machine learning defines the future of AI-based fraud prevention and is the foundation of the top nine ways AI prevents fraud:

- AI is re-defining fraud prevention from relying only on past experiences to taking into account emerging activities, behaviors, and trends in transaction anomalies. By combining supervised learning algorithms trained on historical data with unsupervised learning, digital businesses gain a greater level of acuity and clarity about the relative risk of customers' behaviors.
- **AI makes it possible to detect fraud attacks in real-time versus having to wait six or eight weeks until chargebacks start coming in.** By balancing supervised and unsupervised learning, AI alleviates the need always to play catch-up to online fraud.
- By having an AI-based fraud prevention system do the work of evaluating historical data and anomalies, customer experiences can stay more positive, and the more sophisticated nuanced abuse attacks can be stopped.
- **Provides fraud analysts with real-time risk scores and greater insight into where best to set threshold scores to maximize sales and minimize fraud losses.** Adding in anomaly detection and insights into real-time activity using unsupervised machine learning, fraud analysts can instantly validate or redefine their decision regarding threshold levels, managing risk well.
- **AI enables digital businesses to gain greater control over chargeback rates, decline rates, and operational costs so that business objectives can be achieved.** Digital businesses are relying on the combination of supervised and unsupervised machine learning to attain greater levels of agility, speed, and time-to-market, with AI-based fraud prevention systems being foundational to that effort.
- **Enables digital businesses selling virtual goods, including gaming, to provide a more consistent, high-quality user experience on a 24/7 basis**. AI makes it possible for gamers to buy the coins or tokens they need when they need them to keep playing. AI-based fraud prevention systems make it possible to immediately accept the transactions while still staying within the chargeback thresholds from American Express, MasterCard, VISA, and others.
- **AI reduces the friction customers experience by helping merchants easily approve online purchases and reduce false positives.** AI-based fraud scores like Omniscore reduces false positives, which is a major source of friction with customers. All this translates into fewer manual escalations, declines, and an overall more positive customer experience.
- **Staying in compliance with internal business policies, those from regulatory agencies and agreements with distribution partners is where AI-based fraud prevention is contributing today**. AI-based scoring and fraud prevention are extensively used to keep businesses in compliance.
- **Enables low-margin businesses and product lines to stay profitable by controlling chargebacks levels that have a direct impact on margins.** AI-based approach that incorporates both unsupervised and supervised learning pays off from a gross margin standpoint.

Mengidis et al. [2] also state that including artificial intelligence learning systems in cybersecurity helps prevent attacks in a system. The learning-based system learns from the attackers' actions ad adjusts to protect the information. This factor makes it impossible for attackers to gain access to the data. Having a system that keeps adjusting and learning is one of the attributes that has made the technology very efficient. AI has been able to avoid cyber-attacks using the approaches discussed below. The different techniques ensure the efficiency of AI in cybersecurity.

Applications of AI in Cybersecurity

A. Signature Based Techniques

Signature intrusion detection systems (SIDS) use pattern matching techniques to detect a known attack; these are also referred to as Knowledge-based Detection or Misuse Detection. 6 Matching methods are used in SIDS to locate a previous intrusion triggering an alarm signal whenever an intrusion signature matches one from a previous intrusion existing in the signature database. The most well-known SIDS currently available are Snort, Suricata, NetSTAT and Bro. The method involves AI detecting cyberattacks and malware through the available codes [9]. The database where malware signatures are stored, is called the blacklist. The system detects the attack by comparing the available signatures in the blacklists to the known signature caught in the attack. The signatures are sometimes referred to as the patterns present in the attack, and this could be said to be another form of machine-based learning [3]. Although the method has proven very efficient over the years, it has been seen to be useless in the case of a new attack. The technique fails since the database has no record of the attack [1]. Changing their patterns ensures they can access the data and information before they are detected.

B. Machine Learning

Mengidis et al. [1] discovered that humans always make mistakes when analyzing data or information. The AI technology detects systems, analyses the available records, and detects logs included in the system. This factor ensures that system administrators can change the information accessed to avoid further loss. This factor has led to the analogy that AI closely replaces human analysts. Classification and clustering are great attributes of machine learning systems [8]. They compare the available information and how it should be in the logs. This factor provides detection if there are errors in the system. The regular records are compared to the current ones to identify the infected logs. After an attack has been detected, necessary steps are taken to ensure the attack has been stopped. Clustering involves grouping the available records or information from the system and detecting anomalies. Both of these techniques used in machine learning have proven effective since it is impossible for humans.

C. Network Intrusion Detection

Network attacks are one of cyber security's most used forms of aggression. The raids are conducted through the networks that the organizations or companies use. It is always important to detect attacks through networks. This factor gives the system the advantage of stopping the attack from the web. Stopping attacks from the web is the first step in protecting the information available. This approach has thus been very efficient in preventing

future attacks [9]. The main key attribute and advantage of network intrusion detection systems are that they have five elements that support the full security of such networks. The first key element is how AI systems acquire large sums of information from the network. This factor can be achieved through the AI system's ability to analyze large amounts of data [3]. All the factors help ensure the security of the network has been completed. Stopping an attack from the network gives the organization a higher chance of protecting the information. All the way a network may be compromised is avoided using the AI techniques available.

D. Phishing Attacks

Phishing attack involves some type of social engineering where the attacker sends fraudulent messages tricking the victim into providing his credentials [4]. Attackers could also engage in a mass phishing attack which targets a group of people to directly impact the vulnerable individuals [7]. AI-based Cybersecurity awareness training should therefore be adopted by more organizations, thus reducing the number of global phishing attacks.

The market for artificial intelligence in cybersecurity is also anticipated to grow during the forecast period as a result of the proliferation of 5G technology [5] and the rising demand for cloud-based security solutions among small and medium-sized organizations. In order to secure information, artificial intelligence in cybersecurity is gradually gaining prominence. The market for artificial intelligence in cybersecurity is continuously expanding because end users are anticipated to adopt AI in cybersecurity to solve security problems and recognise new forms of assaults that can occur at any time.

IV. AI CONTRIBUTION IN CYBERSECURITY

While enhancing security, artificial intelligence and machine learning can make it simpler for hackers to break into networks without human assistance. This has the potential to seriously harm any business. In order to want to minimise damages and maintain the viability of a company, getting some sort of security against cybercriminals is of the utmost importance. AI concerns for cybersecurity, as well as advanced tools to thwart attack.

Intelligent attacks

1. **AI-driven malware:** Artificially intelligent malware. In the hands of hackers, AI-driven malware can infect devices faster than ever before, becoming harder to detect, targeting more victims and creating more convincing phishing attacks. [4]
2. **Vulnerable applications:** Weak application security creates an entry point for hackers to infiltrate. Cybercriminals use AI to hide malicious codes within applications, sometimes programming the attack to execute well after the app has been installed. Malicious codes can be present for years before they strike.
3. **Expanded attack surface:** Through hybrid work, employees are spread across distant locations and often access the cloud on their personal devices. This widens the surface area for cyberattack. AI creates an even more evolved threat landscape, expanding what devices and machines can be used to infiltrate systems.

4. **Constant evolution:** Artificial intelligence is always evolving. While the benefits include enhanced threat detection, there are two sides to that coin. Cybercriminals are constantly learning from existing AI tools to develop more advanced attacks and improve their malware. Security must constantly evolve as well.

The use of AI technology has a wide range of effects across numerous industries. These effects cover both this technology's advantages and drawbacks. It is evident from the explanation above that the advantages of AI for cyber security outweigh any drawbacks. More study is being done on artificial intelligence as it continues to develop. This aspect demonstrates the significant technological developments being made in the methods employed to guarantee cybersecurity support. The practises demonstrate the technological influence of cybersecurity measures. The research mentioned above also focuses on various AI-related cyber security limitations. The boundaries demonstrate how individuals have been able to exploit AI to their advantage. This element has caused cybersecurity to be constrained.

V. CONCLUSION

AI and machine learning are redefining every aspect of cybersecurity today. From improving organizations' ability to anticipate and thwart breaches, protecting the proliferating number of threat surfaces with Zero Trust Security frameworks to making passwords obsolete, AI and machine learning are essential to securing the perimeters of any business. Rapid growth of cyber threats and sophistication of cyberattacks require new, more robust, flexible, and scalable methods. In current research, the main targets of AI-based algorithms for cybersecurity are malware detection, network intrusion detection, and phishing and spam detection. Various researches leveraged a combination of different AI techniques, such as ML/DL methods together with bioinspired computation, or different learning methods such as supervised learning together with reinforcement learning. Although the role of AI in solving cyberspace issues is inevitable, some problems related to trust to AI and AI-based threats and attacks would be another concern in cyber environment.

VI. REFERENCES

[1]. Meeraj Farheen Ansari, Bibhu Dash, Pawankumar Sharma, Nikhitha Yathiraju, "The Impact and Limitations of Artificial Intelligence in Cybersecurity: A Literature Review:" Vol. 11, Issue 9, September 2022 DOI: 10.17148/IJARCCE.2022.11912.

[2]. P. Panagiotou, N. Mengidis, et al., Thodora Tsikrika, Stefanos Vrochidis, Ioannis Kompatsiaris, "Host-based Intrusion Detection Using Signature-based and AI-driven Anomaly Detection Methods", Vol. 50, no. 1 (2021): 37-48, ISSN 1314-2119(online).

[3]. Areej Fatima, Tahir Abbas Khan, Tamer Mohamed Abdellatif, Sidra Zulfiqar, Muhammed Asif, Waseem, "Impact and Research Challenges of Penetrating Testing and Vulnerability Assessment on Network Threat", 2023 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2023, pp. 1-8, DOI: 10.1109/ICBATS57792.2023.10111168.

[4]. Katanosh Morovat, Brajendra Panda, "A Survey of Artificial Intelligence in Cybersecurity", 2020 International Conference on Computational Science and Computational Intelligence (CSCI), DOI: 10.1109/CSCI5 1800.2020.00026.

[5]. Meraj Farheen Ansari, Pawan Kumar Sharma, Bibhu Dash, "Prevention of Phishing Attacks Using AI-Based Cybersecurity Awareness Training", International Journal of Smart Sensor and Adhoc Network", Vol 3, Issue 6, March 2022.

[6]. Rammanohar Das and Raghav Sandhane, "Artificial Intelligence in Cybersecurity", ICACSE 2020, Journal of Physics: Conference Series, 1964(2021)042072, DOI: 10.1088/1742-6596/1964/4/042072.

[7]. Zhibo Zhang, Hussam Al Hamadi, Ernesto Damiani, Chan Yeob Yeun, And Fatma Taher, "Explainable Artificial Intelligence Applications in Cyber Security: State-of-the Art in Research", 5th September 2022, IEEE Access, DOI: 10.1109/ACCESS.2022.3204051.

[8]. Navid Ali Khan, Noor Zaman Jhanjhi, Sarfraz Brohi, "Ten Deadly Cyber Security Threats Amid COVID-19 Pandemic", TechRxiv Powered by IEEE – 2020, DOI: 10.3622/techrxiv. 1227879.v1, researchgate.net.

[9]. Bibhu Das, Meraj Farheen Ansari, Pawankumar Sharma, Azad Ali, "Threats and Opportunities with AI-Based Cyber Security Intrusion Detection: A Review", International Journal of Software Engineering and Applications, Vol.13, No. 5, September 2022.

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Leveraging Robotic Process Automation and OpenAI Technologies for Enhanced Business Process Automation

Shilpa Prabhu Patil
Assistant Professor
Department of Computer Science, East Point College of Engineering and Technology, Bengaluru, Karnataka, India

ABSTRACT
Robotic Process Automation (RPA) has revolutionized the automation landscape by allowing businesses to streamline their processes, reduce operational costs, and improve overall efficiency. OpenAI, on the other hand, has emerged as a leading provider of artificial intelligence solutions, offering cutting-edge language models and natural language processing capabilities. This research paper explores the synergy between RPA and OpenAI technologies to create a powerful combination that can further enhance business process automation. The paper delves into the integration of RPA with OpenAI's language models, discussing potential use cases, benefits, challenges, and future directions of this integration.

Keywords - Robotic Process Automation (RPA), OpenAI, NLP, Business Process Automation, AI

I. INTRODUCTION

A. Background
Robotic Process Automation (RPA) has emerged as a game-changing technology in recent years, enabling organizations to automate repetitive and rule-based tasks with software robots. RPA solutions have been widely adopted across industries, significantly improving process efficiency and reducing operational costs. Concurrently, OpenAI has made remarkable strides in the field of artificial intelligence, particularly with the development of advanced language models capable of natural language processing (NLP). OpenAI's language models, such as GPT-3, have demonstrated remarkable language understanding and generation abilities, opening up new possibilities for automation and decision-making. The integration of RPA with OpenAI technologies presents an exciting prospect for businesses seeking to enhance their automation capabilities and harness the power of intelligent automation.

B. Motivation
The motivation behind this research stems from the potential synergies between RPA and OpenAI technologies. By integrating the capabilities of NLP and language processing with RPA, businesses can unlock a new era of automation that is contextually aware, adaptive, and capable of handling unstructured data. The combination of RPA and OpenAI not only promises improved accuracy and efficiency in process automation

but also opens doors to novel applications, such as intelligent document processing, sentiment analysis, and customer service automation. This research aims to explore the possibilities of this integration and provide insights into its real-world implications.

C. Objectives

The primary objectives of this research paper are:

- To investigate the integration of OpenAI language models into RPA workflows and explore how NLP capabilities can augment automation processes.
- To identify and showcase various use cases where RPA and OpenAI integration can bring significant value to different industries and business functions.
- To analyse the benefits and advantages of combining RPA and OpenAI technologies, including improved decision-making, scalability, flexibility, and cost reduction.
- To address the ethical considerations and challenges related to the adoption of AI-driven automation and propose mitigation strategies.
- To outline the implementation challenges that organizations might encounter during RPA-OpenAI integration and suggest strategies to overcome them.
- To provide insights into the future directions of RPA and OpenAI technologies, including advancements, hybrid approaches, and industry-specific applications.
- To present case studies and real-world success stories of organizations that have successfully implemented RPA and OpenAI integration to demonstrate the practical benefits

II. LITERATURE REVIEW

A. Overview of Robotic Process Automation

Robotic Process Automation (RPA) is a technology that utilizes software robots, also known as bots, to automate repetitive, rules-based tasks typically performed by humans. These bots interact with applications and systems just as a human user would, mimicking human actions and following predefined workflows. RPA enables organizations to streamline processes, increase operational efficiency, reduce errors, and free human resources for more strategic tasks.

B. OpenAI Language Models and NLP Capabilities behaviour

OpenAI has been at the forefront of advancing artificial intelligence, particularly in natural language processing. Their language models, such as the Generative Pre-trained Transformer 3 (GPT-3), have shown remarkable language understanding and generation capabilities [1]. GPT-3 can process vast amounts of textual data, understand context, and generate human-like responses, making it a powerful tool for language-related tasks

C. Existing Research on RPA and AI Integration

The integration of AI technologies, including NLP, with RPA has been a subject of increasing interest in the research community. Studies have explored the potential of combining RPA with various AI techniques, including machine learning and natural language processing, to improve automation capabilities and handle unstructured data [2]. Previous research has provided insights into the benefits, challenges, and practical applications of RPA and AI integration

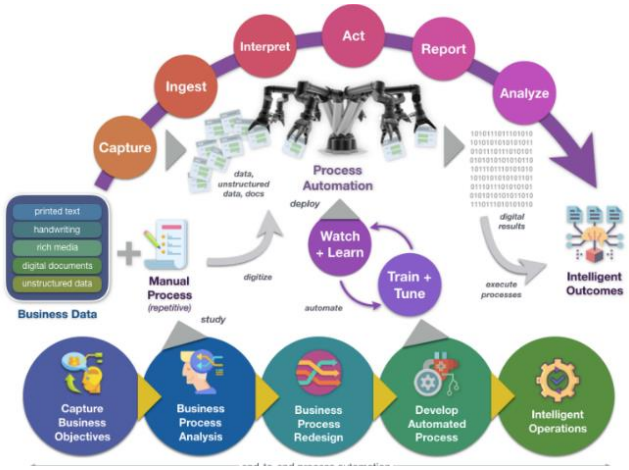


Fig:1 RPA becomes end to end Business Automation

III. RPA AND OPEN AI INTEGRATION

A. Integrating OpenAI Language Models into RPA Workflows

Integrating OpenAI language models into RPA workflows involves leveraging APIs or direct SDK integration to access the language processing capabilities. RPA tools can interact with the language models to comprehend and generate human-like text, making them more contextually aware and adaptable.

B. Natural Language Processing (NLP) Applications in RPA

The integration of NLP in RPA enables a wide range of applications, such as intelligent document processing, sentiment analysis, and customer service automation. Bots can understand and respond to natural language queries, analyse text data, and make informed decisions based on language inputs

C. Challenges and Solutions in RPA-OpenAI Integration

Integrating RPA with OpenAI technologies presents some challenges, including model training, data privacy, and managing unstructured data. Organizations need to address these challenges through proper data handling, model selection, and robust governance frameworks.

IV. USE CASES OF RPA AND OPENAI INTEGRATION

A. Intelligent Document Processing

The combination of RPA and OpenAI's NLP capabilities allows organizations to automate document processing tasks, such as data extraction, content summarization, and document classification. Bots can efficiently process unstructured data from documents and extract relevant information.

B. Customer Service Automation:

By integrating NLP capabilities, RPA-powered chatbots can offer more sophisticated customer service solutions. They can comprehend natural language queries, provide contextually relevant responses, and offer personalized support, enhancing the overall customer experience.

Fig:2 Use cases that can be automated using RPA and OpenAI

C. Sentiment Analysis and Social Media Automation

RPA-OpenAI integration facilitates sentiment analysis on social media platforms, enabling businesses to monitor brand sentiment and customer feedback in real-time

D. Data Extraction and Analysis

RPA and OpenAI integration can streamline data extraction from various sources, such as web pages, emails, and reports. The combined technologies can process and analyse large volumes of unstructured data, enabling data-driven decision-making.

V. BENEFITS AND ADVANTAGES

A. Improved Accuracy and Efficiency

The integration of OpenAI's NLP capabilities into RPA workflows enhances accuracy by enabling bots to understand and process language-based inputs more accurately. This results in more efficient and reliable automation of tasks.

B. Enhanced Decision Making

RPA-OpenAI integration allows organizations to leverage language models for better decision-making. Bots can analyse textual data, identify patterns, and make informed decisions based on the analyzed information.

C. Increased Scalability and Flexibility

The combined technologies offer scalable automation solutions that can handle complex and dynamic business processes. RPA-OpenAI integration provides flexibility to adapt to changing business requirements.

D. Cost Reduction and Resource Optimization

By automating repetitive tasks and leveraging AI capabilities, organizations can optimize resource allocation and reduce operational costs. Human resources can be allocated to more strategic and creative tasks

VI. ETHICAL CONSIDERATIONS

A. Privacy and Data Security

As RPA and OpenAI deal with sensitive data, it is crucial to prioritize privacy and data security. Organizations must implement robust data protection measures to safeguard user information and comply with data privacy regulations [5].

B. Bias and Fairness in AI-driven Automation

The integration of AI technologies must address biases that may arise in AI models, especially when dealing with language-based data. Implementing bias detection and mitigation strategies is essential to ensure fairness and equity in automation processes.

VII. IMPLEMENTATION CHALLENGES AND MITIGATION STRATEGIES

A. Training and Maintaining AI Models

Challenge: AI models require substantial data for training and maintaining them necessitates continuous updates and improvements to adapt to evolving business needs.

Mitigation Strategies: Implement robust data governance practices to ensure data quality and availability. Employ transfer learning to fine-tune pre-trained AI models efficiently and adopt continuous learning approaches to keep the models up-to-date.

B. Integration Complexity

Challenge: Integrating RPA with OpenAI technologies can be complex, requiring seamless communication between systems and ensuring compatibility.

Mitigation Strategies: Utilize APIs and SDKs provided by OpenAI to simplify integration. Adopt a modular approach to development, allowing for flexibility and scalability in integrating AI components with RPA workflows.

C. Regulatory Compliance

Challenge: Implementing AI-driven automation must comply with data protection regulations and industry-specific standards.

Mitigation Strategies: Stay informed about evolving regulations and design systems with privacy and compliance in mind. Conduct regular audits and assessments to ensure adherence to relevant guidelines.

VIII. CONCLUSION

In conclusion, the integration of Robotic Process Automation (RPA) with OpenAI technologies presents a remarkable opportunity to revolutionize business process automation. By combining the power of RPA's rule-based automation with OpenAI's sophisticated language models and natural language processing capabilities, organizations can achieve enhanced efficiency, accuracy, and scalability in their automation endeavours.The research highlights the significant benefits of this synergy, including improved decision-making, seamless data extraction, and superior customer service automation. Moreover, the exploration of ethical considerations underscores the importance of transparency, fairness, and privacy in AI-driven automation.

Looking ahead, the future of business automation lies in advancements in RPA and OpenAI technologies, as well as the emergence of hybrid approaches combining RPA, AI, and machine learning. Real-world case studies further validate the practical value of RPA-OpenAI integration across various industries, demonstrating its potential to shape the automation landscape.

Ultimately, the potential of this symbiotic relationship inspires organizations to embrace intelligent automation, driving innovation, and transforming the way businesses operate in an increasingly AI-driven world.

IX. REFERENCES

[1]. https://www.researchgate.net/publication/369216541_Intelligent_Automation_Framework_Using_AI_and_RPA_An_Introduction

[2]. Mohanty S, Vyas S (2018) It operations and ai. In: How to compete in the age of artificial intelligence. Apress, Berkeley, CA, pp 173–187

[3]. Kung, T.H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Tseng, V.: Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLoS digital health, 2(2), e0000198. (2023)

[4]. https://www.fintechfusion.io/p/chatgpt-a-game-changer-in-finance Accessed 15 Feb 2023

[5]. Ventayen and Randy Joy Magno, "OpenAI ChatGPT Generated Results: Similarity Index of Artificial Intelligence-Based Contents", Available at SSRN 4332664, 2023.

[6]. CS, Diwaan Chandar and Swasthika AK, "Analyzing and Experimenting Open Source OCR Engines in RPA with Levenshtein Distance Algorithm", International Research Journal on Advanced Science Hub, vol. 2, pp. 98-104, 2020.

[7]. Jorge Ribeiro, Rui Lima, Tiago Eckhardt and Sara Paiva, "Robotic process automation and artificial intelligence in industry 4.0–a literature review", Procedia Computer Science, vol. 181, pp. 51-58, 2021

'National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Smart Agricultural System for Pests Detection Using IOT

Pavan Kumar R B[1], Pooja P P[1]
[1]Assistant Professor
Department of Computer Science, East Point College of Engineering and Technology, Bengaluru, Karnataka, India

ABSTRACT
Artificial intelligence and image recognition technologies are combined with environmental sensors and the Internet of Things (IoT) for pest identification. In India, about 70% of population depends upon farming and one third of the nation's capital comes from farming. Currently all over the world, it is found that around 50% of the farm produce never reach the end consumer due to wastage. Smart farming is an emerging concept, because IoT sensors capable of providing information about soil pH, soil moisture, temperature, humidity. The proposed design has been developed with acoustic sensor and PIR sensor. To achieve automatic recognition of agricultural pests, we developed a pest recognition system based on image processing technique. The image segmentation technique is used to detect the presence of pests in leaf images.
Index Terms: Artificial intelligence (AI), Internet of Things (IoT), Pest Monitoring, Acoustic, PIR sensor.

I. INTRODUCTION

The information you provided highlights the critical importance of managing pests and diseases in crop production and the significant economic impact they can have on farmers. Early detection of pest hotspots and timely application of pesticides are crucial for mitigating losses and optimizing crop yields.

In this context, the integration of deep learning and image recognition technologies in agricultural production offers promising solutions. Deep learning models, especially convolution neural networks (CNNs), have shown remarkable efficiency in diagnosing crop diseases from images. These models can be trained on a vast dataset of images of healthy and diseased plants to learn to recognize patterns and characteristics associated with different diseases and pests.

By deploying deep learning-based image recognition systems in the field, farmers can quickly and accurately detect the presence of diseases and pests on their crops. This allows for timely intervention, such as targeted pesticide application, reducing the need for excessive and indiscriminate pesticide use, which can be harmful to the environment and costly for the farmers.

The fact that the deep learning model achieved an 83.5% recognition rate is encouraging, but it's essential to continually improve and fine-tune the model to achieve even higher accuracy levels. This can be done through regular updates and retraining with new and diverse datasets to account for variations in environmental conditions and the appearance of different disease and pest types.

By reducing the dependence on plant protection technicians, the image recognition system empowers farmers to address crop issues promptly and make informed decisions regarding crop management. This not only improves productivity but also helps farmers optimize the use of resources and reduce overall costs. Furthermore, the ability of the deep learning model to monitor, diagnose, and prevent crop growth issues in real-time can have a transformative effect on agriculture. Early detection and intervention can save crops from substantial damage and increase the likelihood of achieving healthy yields.

As more farmers adopt these technologies, especially in regions where agriculture is a significant source of income, there is potential for substantial positive economic and environmental impacts. By minimizing crop losses and optimizing resource usage, sustainable agricultural practices can be promoted, leading to better livelihoods for farmers and ensuring food security for the population. Overall, the combination of deep learning, image recognition, and IoT technologies has the potential to revolutionize agriculture, making it more efficient, sustainable, and resilient to pests and diseases. Continued research, development, and implementation of such solutions are essential for the future of global food production.

II. EXISTING SYSTEM

The challenges in early detection of diseases and pests in agriculture can have significant implications for crop productivity and overall agricultural sustainability. The current methods, such as pheromone-based glue traps, have limitations in detecting pests at low densities and may not provide timely detection to prevent significant damage. Manual pest identification, relying on the expertise of trained individuals, is time-consuming and requires continuous monitoring. Moreover, human error and subjective judgments can lead to inaccuracies in pest identification, which can result in delayed or inadequate responses to pest outbreaks. The absence of a pest monitoring system for Ginger highlights the need for specialized solutions tailored to different crops. Generalized approaches may not effectively address the unique challenges posed by specific crops and pests.

The disadvantages of the existing system, as mentioned, include:

Delayed Detection: The current system only identifies pests after they have already affected the plants. Early detection is crucial to prevent the rapid spread of pests and to implement timely control measures.

Increased Pesticide Usage: In the absence of timely detection, farmers may resort to increased pesticide usage as a reactive measure to control pests. This can lead to the overuse of chemicals, negatively impacting the quality of the crop produced, as well as posing environmental and health risks.

To address these issues, the development of advanced pest monitoring and detection systems using modern technologies, such as AI-powered image recognition and IoT, can significantly improve the efficiency and accuracy of pest identification. Here are some potential solutions:

1. Automated Pest Detection: Utilizing image processing techniques and AI algorithms, automated systems can analyze images of crops to identify pests and diseases accurately. These systems can quickly scan large areas and provide real-time alerts when pests are detected, enabling farmers to take immediate action.
2. IoT-Enabled Sensors: Deploying IoT sensors in the fields can provide valuable data on environmental conditions, such as temperature, humidity, and soil moisture, which can influence pest behavior. Integrating this data with pest detection systems can improve the accuracy of predictions and aid in early pest detection.

3. Crop-Specific Solutions: Tailoring pest detection and monitoring systems to specific crops, like Ginger in this case, can account for the unique pest profiles and challenges faced by each crop, leading to more effective pest management.
4. Precision Pest Control: Early and accurate pest detection allows for targeted and precise application of pesticides, reducing overall pesticide usage and its negative impacts on crop quality and the environment.

By leveraging technology and data-driven approaches, the agriculture industry can move towards proactive and sustainable pest management practices. Early detection and precise control of pests can significantly improve crop yields, reduce costs, and promote environmental sustainability.

III. PROPOSED SYSTEM

Our proposed system for automatic pest detection and classification in agriculture sounds promising and has several advantages. By utilizing modern technologies like Raspberry Pi, IR sensors, and ultrasonic sensors, you can create an efficient and user-friendly solution for farmers to monitor pest infestations and plant growth. Let's take a closer look at the advantages you mentioned:

1. Early Detection of Pests: Early detection is crucial in pest management as it allows farmers to take timely action to control and prevent further infestations. By identifying pests at an early stage, farmers can minimize crop damage and reduce the cost and amount of pesticides used for harvesting.
2. Reduced Pesticide Usage: With early detection of pests, farmers can apply targeted and localized pesticide treatments only where they are needed, rather than blanket spraying the entire crop. This targeted approach helps minimize the environmental impact and potential harm to non-target organisms.
3. Easy to Use: Your proposed system seems user-friendly and accessible to ordinary farmers. The integration of sensors with Raspberry Pi and the option to receive pest detection alerts via email and view them on PC or mobile devices make it convenient for farmers to monitor their crops remotely.
4. Enhanced Crop Management: The combination of pest detection and the ability to measure plant growth using ultrasonic sensors can provide valuable insights into the overall health and development of crops. This information can help farmers optimize their agricultural practices and make informed decisions for better crop management.
5. Cost-Effectiveness: By reducing pesticide usage and increasing the efficiency of pest detection and control, your proposed system can potentially save farmers money in the long run. Improved crop yields and reduced crop losses contribute to increased profitability in agriculture.
6. Technology Adoption: Introducing modern technologies like IoT sensors and Raspberry Pi in agriculture can encourage technology adoption among farmers, leading to more sustainable and efficient farming practices.

While the proposed system holds great potential, it's essential to thoroughly test and validate its accuracy and reliability in real-world farming scenarios. Additionally, considering factors such as weather conditions, pest species diversity, and variations in plant types will be crucial for the success of the system.

Overall, the proposed system can contribute significantly to improving pest management practices in agriculture, leading to enhanced crop yields, reduced pesticide usage, and better resource management.

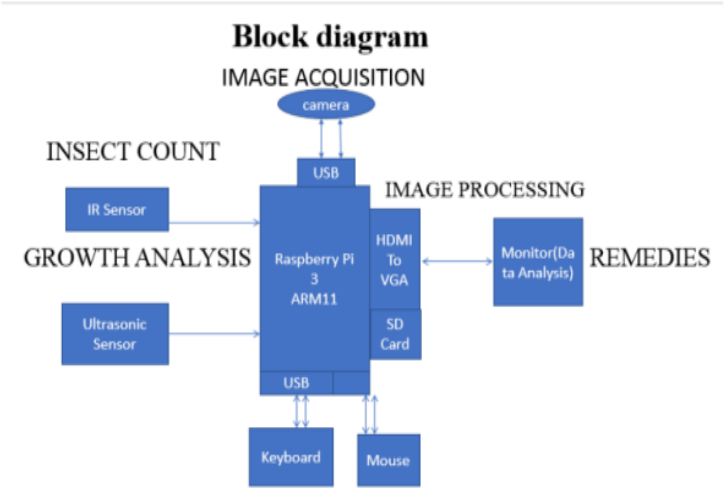


Image Acquisition:

The first stage of any imaginative and prescient device is the photo acquisition stage. After the photo has been received, Various techniques of processing can be carried out to the photograph to perform the various extraordinary vision obligations required today. Virtual photo acquisition is the creation of photographic pictures, such as of a physical scene or of the indoors Shape of an object. A digital picture may be created immediately from a bodily scene by way of a digicam or comparable tool.

Image Processing and classification:

The image processing is a method to perform some operations on an image, with the intention to get a stronger image. It could be done with the aid of the usage of OpenCV(Open Source Computer Vision Library) and Python. It's far a type of signal processing wherein input is an photo and output can be picture or function/capabilities related to that image.

Image processing basically consists of the following three steps:

- uploading the images through images acquisition equipment
- studying and manipulating the image
- Output

Image class is possibly the most crucial part of virtual picture analysis. The reason of the Class process is to categorized all pixels in virtual pix.The goal of image category is to perceive and portray, as a unique gray stage (or shade), the functions occurring in a picture in terms of the item or form of landcover these functions sincerely constitute at the floor.

Detection and Remedies:

Once image processing and type is done, the machine intimates to the user if it is a bad insect which influences the crop. Additionally, it proposes the pesticides which may be used for the eradication of such pests.The intimation message is dispatched to the user by e-mail so they can be able to view the message in both mobile and personal computer.

Growth Analysis:

Ultrasonic sensors are a device that could degree the distance to an object via the usage of sound waves. It measures the distance via sending out a valid wave at a specific frequency and listening for that sound wave to bounce back. By recording the elapsed time among the sound wave being generated and the sound wave bouncing returned, it's miles feasible to calculate the distance among the sonar sensors and the object. It's miles crucial to understand that some gadgets might not be detected through ultrasonic sensors. That is due to the

fact some items are fashioned or located in the sort of way that the sound wave bounces off the item, however are deflected far from the Ultrasonic sensor. It is also viable for the object to be too small to reflect enough of the sound wave back to the sensor to be detected.

Insect Count:

An infrared sensor is an electronic device that's used to feel certain traits of its Environment by both emitting and/or detecting infrared radiation. Infrared sensors are also capable of measuring the heat being emitted by an object and detecting motion. IR Sensors paintings via the usage of a selected light sensor to hit upon a pick out mild wavelength in the Infra-red (IR) spectrum. By using an LED which produces mild at the equal wavelength as what the sensor is searching out, you could observe the intensity of the obtained light. Whilst an item is near the sensor, the mild from the LED bounces off the item and into the light sensor. This effects in a massive bounce inside the intensity, which we already recognize can be detected the use of a threshold.

IV. CONCLUSION

In this paper, the automatic detection is used for image processing. One of a kind picture processing techniques have been used to stumble on and extract the pest inside the captured picture. By means of this approach the pest may be diagnosed in its early degree. For this reason, we are able to lessen the use of pesticides has been multiplied which in flip reduces the great of crop produced. The correct disease detection and category of the plant leaf image could be very critical for the successful cultivation of cropping and this could be done the use of photograph processing. This paper discussed numerous techniques to phase the sickness of the plant. This paper mentioned class strategies to extract the functions of inflamed leaf and the class of plant illnesses through Raspberry Pi.

V. REFERENCES

[1]. Plant Chili Disease Detection Using The RGB Color Model By Zulkifli Bin Husin, Ali Yeon Bin Md Shakaff, Abdul Hallis Bin Abdul Aziz, And Rohani Binti S Mohamed Farook, Research Notes in Information Science (RNIS) Volume13, May 2013.

[2]. Image Processing Approach For Grading And Identification Of Diseases On Pomegranate Fruit By S. Gaikwad, K. J. Karande /(IJCSIT) International Journal of Computer Science and Information Technologies Vol. 7 (2), 519522, 2016.

[3]. A Review For Agricultural Plant Diseases Detection Using Different Techniques By Mr.N.P.Kumbhar, Dr.Mrs.S.B.Patil , International journal of Electrical and Electronics Engineering(IJEEE) vol no.9,Issue no.1,January-june 2017.

[4]. Disease Detection And Diagnosis On Plant Using Image Processing By Mr.KhushalKhairnar,Mr.Rahul Dagade. Volume 108 – No. 13, December 2014.

[5]. Rice Disease Identification Using Pattern Recognition, Proceedings by Santanu Phadikar And Jaya Sil, 11th International Conference On Computer And Information Technology (ICCIT 2008) 25-27 December, 2008, Khulna, Bangladesh.. International Journal of Scientific Research and Review Volume 7, Issue 3, 2018 ISSN NO: 2279-543X 124 http://dynamicpublisher.org/

[6]. Ai, Yong, et al. “Research on Recognition Model of Crop Diseases and Insect Pests Based on Deep Learning in Harsh Environments.” IEEE Access, vol. 8, 2020, pp. 171686–93, doi:10.1109/access.2020.3025325.

[7]. Brunelli, Davide, et al. E Nergy N Eutral M Achine L Earning B Ased I o T D Evice for P Est D Etection in P Recision A Griculture. no. December 2019, 2020, pp. 10–13.

[8]. Chen, Ching Ju, et al. “An AIoT Based Smart Agricultural System for Pests Detection.” IEEE Access, vol. 8, 2020, pp. 180750–61, doi:10.1109/ACCESS.2020.3024891.

EAST POINT
GROUP OF INSTITUTIONS

National Conference on Recent Advances of Computational Intelligence Techniques in Science, Engineering and Technology

Organized by

Department of Computer Science and Engineering,
East Point College of Engineering and Technology,
Bangalore, Karnataka, India

Publisher

Technoscience Academy

Website : www.technoscienceacademy.com
Email: info@technoscienceacademy.com

Email : editor@ijsrcseit.com, ijsrcseit@gmail.com