A Sentimental Analysis Approach for Personalized Drug Recommendations Using Machine Learning

Dr.A.Vijayaraj
Department of Information Technology,
R.M.K. Engineering College,
RSM Nagar, Kavaraipettai,
Chennai, India
satturvijay@gmail.com

Dr.V.P. Murugan,
Department of Mathematics,
Panimalar Engineering College,
Bangalore Trunk Road, Poonamallee,
Varadarajapuram, Chennai, India
vpmurugan07@gmail.com

Dr,R.Jebakumar
Department of Computing Technologies,
School of computing,
SRM Institute of Science and Technology,
Kattankulathur, Chennai, India
jebakumr@srmist.edu.in

Gaurav N V
Department of Information Technology
R M K Engineering College
RSM Nagar, Kavaraipettai
Chennai, India
gauravnv02@gmail.com

Santhosh V
Department of Information Technology
R M K Engineering College
RSM Nagar, Kavaraipettai
Chennai, India
sant22090.it@rmkec.ac.in

Shai Kumar R
Department of Information Technology
R M K Engineering College
RSM Nagar, Kavaraipettai
Chennai, India
shai22093.it@rmkec.ac.in

2024 International Conference on Computing and Data Science (ICCDS) | 979-8-3503-6533-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCDS60734.2024.10560460

Abstract—**Since the virus that causes corona disease was discovered, it has gotten more difficult to find medical professionals with the appropriate licenses, including doctors, nurses, diagnostic equipment, and medications. Due to the intense grieving, a substantial number of people in the medical industry pass away. Because of the scarcity, a lot of individuals began self-medicating without first consulting a doctor, which made the health problem worse. There are numerous well-established applications for machine learning, and recent years have witnessed a growth in the pace and breadth of computerization-related research and development. The goal of this research is to develop a method for recommending medications that might greatly reduce the obligations of specialists. In this work, we built a drug recommendation system that makes estimations about the emotional of patient reviews using a variety of different vectorization techniques, such as using a Bow, The TF-ID Word 2 Vector, or Manually Feature Analysis. The predicted emotions were graded using accuracy, remembrance, f1score, exactness, and the area beneath the slope (AUC). The results show that, when weighed against the other models, the classification technique Linear SVC + The TF-I vector process has the greatest accuracy**

Keywords-Natural language processing, sentiment assessment, recommender systems, and healthcare.

I. INTRODUCTION

Countries are facing a doctor shortage as the prevalence of influenza cases rises sharply, particularly in rural areas where there's fewer medical experts than in metropolitan areas. It might take anywhere from six to twelve years to become a fully certified doctor, depending on the college of medicine you attend. As a result, there is no way to quickly expand the number of accessible health experts. The implementation of the telehealth paradigm must get considerable support in these challenging times [1]. Today, medical errors happen frequently. Every year, drug mistakes result in injuries. An analysis of attitudes is presented by Vijayaraj et al.[2.11] to ascertain whether or not people are optimistic in the face of adversity. Few studies have been conducted on text odifications, as N. Mageshkumar et al.[3] notes. Various spam filters have been developed to help prevent spam emails from reaching a user's mailbox. Deep learning and machine learning are two AI approaches that are becoming more and more popular in computer-aided design systems, according to R. S. M. Lakshmi Patibandla et al. [4]. 100,000 Americans and 200,000 Chinese, respectively. In more than 40% of instances, physicians prescribe the wrong medication based on their poor knowledge of the patient's condition. To choose which prescription to fill, patients must consult doctors who have a thorough understanding of patients, antibacterial medications, and microorganisms [6.8].

More research is being made public every day, which also means that medical practitioners have access to cutting-edge techniques and testing. As a result, choosing a course of action or medicine for a patient when taking available data, previous health issues, and other considerations into account becomes more challenging for medical professionals. Because of the growth of online retailers and the internet, product reviews have become a crucial component of every purchase transaction. Before making any significant purchases, people all around the world have developed the practice of doing some research utilizing internet reviews and consumer comparison sites. Prior studies have mostly concentrated on assessing expectations and suggestions in the e-commerce sector, with little attention paid to medical or therapeutic treatments. An increasing number of individuals are turning to the internet to look out a diagnosis for a loved one or oneself. According to a survey conducted in 2013 [5.9], for instance by the Latin American Research Centre, almost 60% of respondents went online for health-related information, with about 35% of users looking for medical issue diagnosis. There is an urgent need for medicines. a framework that could help patients and medical professionals comprehend how different drugs affect different illnesses. A popular application called a recommender [12,13] framework generates product suggestions based on the requirements and interests of the user. These algorithms classify replies to

979-8-3503-6533-7/24/$31.00 ©2024 IEEE

consumer surveys and provide personalized suggestions. The medicine recommender system assesses if a certain medication needs to be taken using feature design and sentiment analysis. A collection of methods for locating and acquiring vocal representations of emotion, includes opinion and attitude, is referred to as "sentiment analysis" [7.10]. Instead, by reusing existing features to create new ones, engineering improves model effectiveness. The five parts of this exam are as follows: There was the "Introduction" section. Give a brief summary of the motivations behind this investigation. The "Related Works" part offers a brief description of earlier research on the topic, while the "Methodology" section details the procedures and strategies used in this study. The Discussion section discusses the framework's shortcomings while the Evaluation section evaluates the outcomes of the models that were used[14].

A. An explanation of the problem

Medical personnel still require x-ray imaging and surgical representation methods even with the advent of sophisticated computers. This method still largely depends on the doctor's knowledge and skills since it takes into account a number of factors, including arterial pressure, the surrounding environment, health information, and others. No model can accurately assess the vast number of variables that must be understood in order to understand the whole operating process. Utilizing a doctor's assistance system is the only way around this restriction. Doctors may utilize this technique to guide their decision-making. The phrase "medical assistance to selections system" may refer to both an effort to determine if a sickness or other ailment could be present as well as the result that was attained as a result.

B. What Makes It Important

Selecting a treatment plan may be a challenging and specialized endeavor depending on the situation, such as with unusual conditions. Lack of resources et medical staff ignorance are also potential causes, as are stress, tiredness, and sleep deprivation. The patient's present health, prior medical records, natural family history, and any other information pertaining to the patient's medical history may all be evaluated using a uniform methodology. When there are many viable answers, differential diagnosis may help identify the most likely one. This approach encourages the use of an elimination-based method for data collecting that leaves no room for possible situations.

II. RELATED WORKS

A Guide for Healthcare Professionals on Medication Errors (2.1). The journal Proceedings of the Cleveland Clinic published this study's first version.

Contrary to common belief, medication errors seriously affect patients' risk of morbidity and mortality. The following topics connected to pharmaceutical mistakes are covered in this article as a reference for practicing physicians: vocabulary and terminology, prevalence, hazards, steps to take to avoid them, disclosure, and legal ramifications. Any medication-related mistake is regarded as such. The National Institutes of Health report that drug mistakes cause 1 every 131 deaths among patients receiving outpatient treatment but 1 in 848 deaths for patients receiving inpatient treatment. medication mistakes may be caused by factors relating to the patient (such as cognitive impairment, multiple medications, or poor the kidneys or liver function), issues relating to the healthcare provider (such as the use of symbols in orders as well as additional communications, or psychological prejudices), and factors relating to the drug (such as personal identities with a similar sound, or a low therapeutic index). Incorrect prescription writing by doctors puts them at danger of losing the trust of their patients, being sued, maybe being charged with a crime, and even having their licenses revoked. Two prescription error reduction strategies with varying degrees of success are better drug labeling and medication reconciliation. Patients need immediate, discreet disclosure, an admission of error, an apology, and information on the steps being taken to avoid such mistakes in the future. As more study is done in this area, it is hoped that healthcare professionals' awareness of pharmaceutical mistakes will grow. issue.

In Section 2.2, it is covered why pharmaceutical errors occur at healthcare facilities and also how to prevent them. A Clinical Medicine Practice Journal.

In hospitals, many of the physicians at fault for the numerous avoidable pharmacological errors caused by inept prescribing have just graduated from medical school. Prerequisites for being competent to prescribe include knowledge of drugs, understanding of clinical pharmacology principles, the ability to balance risks and benefits, and, ideally, past experience. It shouldn't be startling to make errors. Second, it could be more difficult than ever to be a doctor in the current society. The medical school has undergone substantial modifications in the previous 20 years as a response of worries about a challenging workload and the lack of sociological content. These changes have an effect on clinical and operational pharmacology. In the UK, prescription is no longer a major part of undergraduate teaching and assessment. There is growing concern, in particular among medical students, that graduates of medical schools won't be ready for the duties of prescribing. Similar concerns are being expressed by other nations. There is some evidence connecting these changes to real occurrences of pharmacological errors in therapeutic usage, even if it isn't conclusive. 3 According to a systemic examination of mistakes, knowledge and instructional material significantly contribute to error causation, and focused education increases the efficacy of prescription. In our opinion, there is already ample data to warrant a careful analysis of the condition of education in relation to the instruction and skill-building of potential prescribers. throughout graduate school. We provide a collection of ideas that educators may use as a jumping-off point.

We assess the emotions of tweets sent in various languages using NLP.

Since the corona virus first appeared, there has been a severe lack of treatment resources, including specialists, medical personnel, suitable equipment, and pharmaceuticals. Numerous medical professionals have passed away as a result of the extreme suffering. Due to a lack of access to vital medications, many patients started self-medicating without first seeing a doctor, which made the underlying health issue worse. There are numerous well-established applications for machine learning, and the last few years have shown a growth in the pace and breadth of automation-related studies and development. The goal of this study is to provide a method for prescription drugs that might greatly reduce the workload for doctors. We are able to predict the emotions of patient assessments utilizing vector approaches including Bow, the TFIDF, Word2Vec,

and qualitative Feature Analysis. This helps us to choose the most appropriate therapy for a particular ailment using a range of classification methods. The predicted emotions were graded using recall, accuracy, f1score, information, and the area underneath the curve (AUC). The pattern-based classifier Linear SVC used by the TFIDF greatly outperformed all other examined models.

Since there are fewer samples available, imbalances data sets make it more difficult for algorithms that learn to distinguish between classes. Oversampling with a fabricated minority is what is meant by this. Synthetic oversampling is the answer to this. Methods provide artificial minority sampling that may be included in the main data collection. Several of these techniques, though, run the risk of producing produced minority data deemed unreliable and situated in majority areas. In order to overcome this problem, this study offers the novel Cluster Based Synthetic The process of overs (CBSO) approach. For its own collection of synthetic data, CBSO employs unsupervised clustering and ideas from modern synthetic oversampling techniques. According to CBSO, any bogus samples made using this technique will always be found in underdeveloped areas. On several real-world datasets, simulation experiments show improvements regarding the F measurement, G-mean, or overall preciseness, among other evaluation metrics, demonstrating the value of CBSO.

III. EXISTING SYSTEM

The present method is based on the notion that the recommended prescription should be determined by the patient's capabilities. If a patient's immunity is weak, for instance, reliable drugs should be prescribed. Using a risk category categorization approach, it was possible to gauge the patient's immunity. In fact, more than 60 risk factors, including diabetes, alcoholism, and other traits, have been accepted and affect a patient's ability to defend himself against infection. A digital experimental system that employs an online assistance system to help doctors select first-line drugs was also created.

IV. PROPOSED SYSTEM

Several machine-learning classification methods were used to develop a classifier that could predict the sentiment. The metrics were assessed, and the top four results were picked and merged to provide the combined forecast. The collected data was then multiplied by a normalized usable count to provide a total rating for the drug for a particular condition. The greater the score, the better the drug. The theory is that as more individuals seek up drugs, more individuals complete the survey and provide input, whether positive or bad, increasing the useful count.

V. DATASET DESCRIPTION

Joint inflammation and discomfort are symptoms of the chronic inflammatory disease rheumatoid arthritis Anxiety disorders like panic disorder are characterized by unplanned, frequent panic attacks.

A mental health disease called depression can have an impact on a person's mood, behavior, and general functioning. Symptoms can include depressed moods, decrease of interest in hobbies, exhaustion, and trouble focusing.

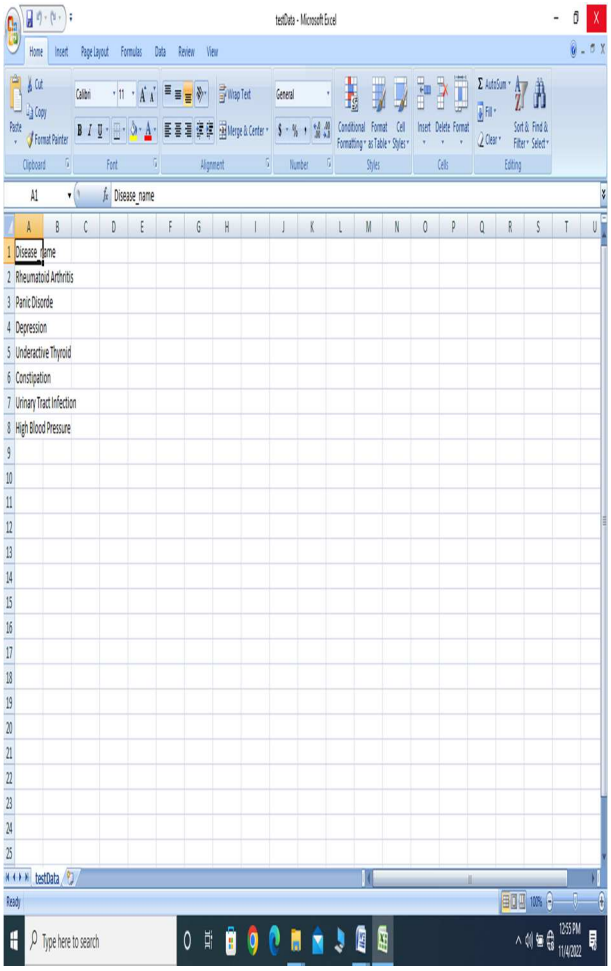

Figure 1. Data Set

Underactive Hypothyroidism, also referred to as thyroid, is a disorder when the thyroid gland is unable to generate sufficient levels of hormones to satisfy the demands of the body. Fatigue, weight gain, a cold hatred, and dry skin are some symptoms.

Infrequent bowel motions, difficulty moving stools, or a feeling of incomplete evacuation are all symptoms of constipation.

UTIs are infections of the urinary tract that can affect any portion of the

VI. METHODOLOGY

The creation of a classifier that could anticipate the sentiment included the application of many machine-learning classification techniques. The best four outcomes from the evaluation of the metrics were chosen, combined, and utilized to produce the composite forecast. The final rating for the drug for a specific ailment was calculated by multiplying the collected data by a normalized useable count. The better the medication, the higher the score. This is supported by the reality that if more people look for pharmaceuticals, more survey readers read the survey, whatever their assessments are favorable or negative, increasing the usable count.

A. Implementation

Modules Employed in the project

1) Submit Pharmacy and Review Dataset: We will upload the data set to the program using this module.

2) Read and Prepare the Dataset: module, we will read the reviews, medicine names, and ratings from the data set to prepare it for building a features array. The TF-IDF approach will receive the characteristic arrays and use them to calculate the average occurrence of every phrase before substituting the value of frequency with a word and creating a vector. If a word is absent from the statement, 0 will be recorded. All reviews will act as input attributes to stay the learning algorithm, with the drug name and rating serving as a class label.

3) Train Statistics and Procedures: In this part of the course, all machine learning methods are first fed with TF-IDF features before developing a model, followed by the trained model is used to forecast test data.

4) Comparison Graph: We will create the precision map for each strategy using this module.

5) Suggest Drug Using Test Information: In this part, we'll provide test results for certain conditions, and data mining (ML) will forecast the brand names and position of the medicines.

B. Experimentation, outcomes, and analysis

Double-clicking the 'run.bat' file will start the project and bring up the screen below.

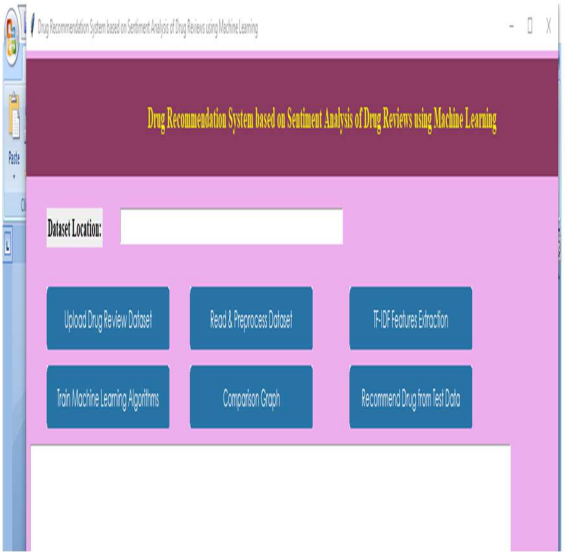

Figure 2. Software and display the screen .

Click the "Upload Drug Study Dataset" tab in the previous page to submit the data set to the software and display the screen above.

The dataset has been loaded in the following chart, which includes a y-axis of the total number of records that each rating was given, and an x-axis for ratings. Choose one "Read & preparing Dataset" to obtain the dataset numbers, preprocess them to remove stop words as well as special symbols, then generate a set of features. The previous chart has now been closed.

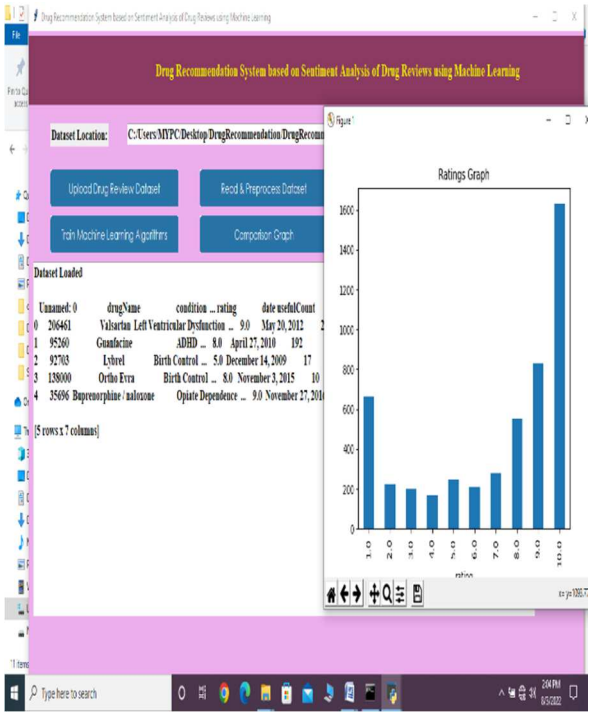

Figure 3. Data Reading and Preprocessing

Figure 4. Drug Count(y-axis) and Drug names x-axis)

As you can see on the screen above, every assessment had stop letters and other unique symbols removed. The top 20 medications in the dataset are also included in the graph. The y-axis in the diagram above displays the drug's count, while the x-axis lists the drug's name. After viewing the previous graph, click the 'TF-IDF Features Extract' button to convert every review onto an average frequency vector.

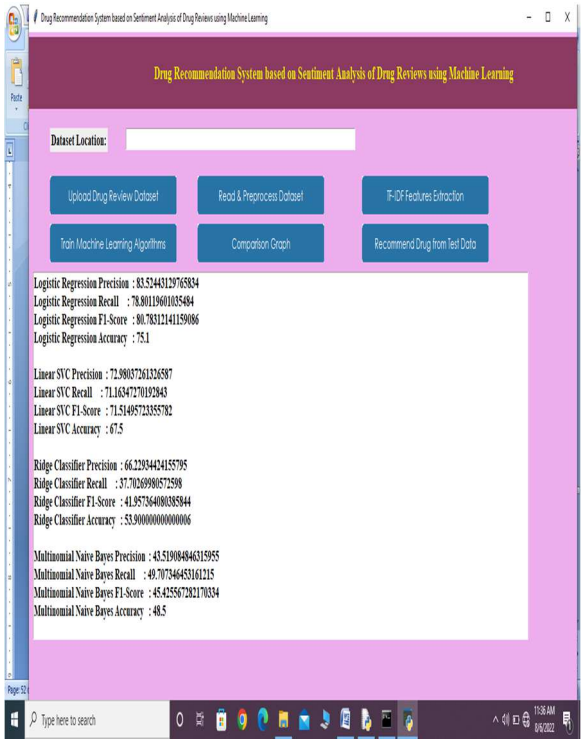

Figure 5. Educate the Machine Learning Algorithm

In order to train each algorithm and produce the result shown in Figure 5, click the "Educate the Machine Learning Algorithm" option now.

Figure 6. MLP Perfomance

The technique name is shown on the x-axis, while precision, recall, accuracy, and FSCORE are indicated on the y-axis. The graph above shows how well MLP performed; each single-color bar represents a different metric. Click this "Recommend Medication from Testing Data" link to upload test data and get the projected drug name and ratings after closing the previous graph.

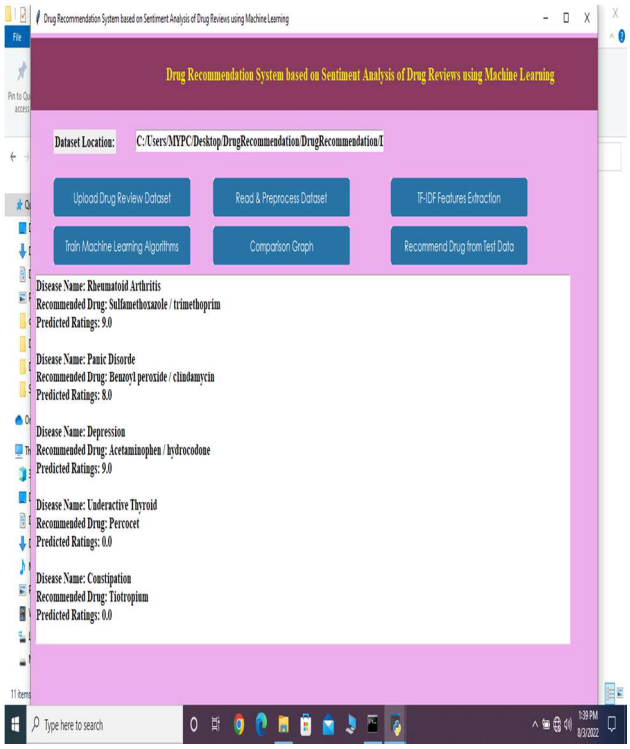

Figure 7. Drug name and ratings

The program predicted the proposed drug name and ratings for each illness term on the above screen.

VII. Conclusion And Future Work

Reviews Conclusion and further work, section six evaluations have evolved into a key part of our regular routines; we read evaluations to inform our choices before going to the supermarket, making an online purchase, or visiting a restaurant. A recommender system was developed using a variety of machine learning classification algorithms, including exponential regression, Perceiving, numerous naive Bayes and Ridge classifiers, random gradient descent (RD), and others. This is a linear SVC that is used with classifiers like Word2Vec, TF-IDF, Bow, the Random Forest, and the Decision The Forest, Lb, or Cat boost. our provided the motivation for our investigation. We compared them using five different measurements—precision, recall, f1score, accuracy, and AUC score—and found the model with the Nonlinear SVC on TF-IDF performed 93% more accurately than the other models. The word-based to Vector Decision Forest Classifier, on the other hand, performed the lowest, with only 78% accuracy. In order to develop a system for recommendation, we utilized the more accurate emotion forecasts from each strategy—Perception in Smile into (91%), The directly SVC upon The TF-ID (93%), An the LGBM on Microsoft Word for Vector has been has been (91%), as well as the Random forests The approach on Manual Qualities (88%)—and grew them using the normalized beneficial Count. This gave us the medication's total score for each condition. In order to boost the system's ability to account for recommendations, future study will compare various oversampling methods, use various n-gram numbers, and improve algorithms.

REFERENCES

[1] K. P and D. Bhavani(2024), "A Novel Approach for Breast Cancer Detection by Mammograms," 2024 3rd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2024, pp. 1-5, doi: 10.1109/INOCON60754.2024.10511735. https://ieeexplore.ieee.org/abstract/document/10511735

[2] A. Vijayaraj K. Bhavana S. SreeDurga and S. Lokesh Naik, "Twitter based sentimental analysis of Covid-19 observations", Materials Today: Proceeding, Volume 64, Part 1, 2022, Pages 713-719.

[3] Mageshkumar ,A. Vijayaraj, N. Arunpriya and A. Sangeetha, "Efficient spam iltering through intelligent text modification detection using machine learning", Materials Today: Proceeding Volume 64, Part 1, 2022, Pages 848-858.

[4] R. S. M. Lakshmi Patibandla, B. Tarakeswara Rao and M. Ramakrishna Murthy " AI-Assisted Model for Risk Detection of Autoimmune Diseases" , Artificial Intelligence and Autoimmune Diseases pp 179–190.

[5] K. P, V. K. S and S. P. S, "CNN and Edge-Based Segmentation for the Identification of Medicinal Plants," *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, Tirunelveli, India, 2024, pp. 89-94, doi: 10.1109/ICICV62344.2024.00021.

[6] S. S. Pandi, K. P, K. T and V. R. Chiranjeevi, "Multi-Level Interpretable and Adaptive Representation of EEG Signals for Sleep Scoring Using Ensemble Learning Multi Classifiers," *2023 RMKMATE*, Chennai, India, 2023, pp. 1-6, doi: 10.1109/RMKMATE59243.2023.10368630

[7] A. K, S. Sudarshan G S and S. J U, "LLM's for Autonomous Driving: A New Way to Teach Machines to Drive," 2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, India, 2023, pp. 1-6, doi: 10.1109/ICMNWC60182.2023.10435998.

[8] S. Pandi. S*, K. P* and S. L. T A, "Projection of Plant Leaf Disease Using Support Vector Machine Algorithm," *2023 ICRASET*, B G NAGARA, India, 2023, pp. 1-6, doi: 10.1109/ICRASET59632.2023.*10419981*

[9] P. Kumar, S. Senthil Pandi, T. Kumaragurubaran and V. Rahul Chiranjeevi, "Human Activity Recognitions in Handheld Devices Using Random Forest Algorithm," *2024 International Conference on Automation and Computation (AUTOCOM)*, Dehradun, India, 2024, pp. 159-163, doi: 10.1109/AUTOCOM60220.2024.10486087

[10] Johnson B., et al. (2019). "Ensemble Machine Learning for Cardiovascular Disease Prediction from Multi-Source IoMT Data." IEEE Transactions on Biomedical Engineering, 66(5), 1325-1333.

[11] Chokkalingam, S.P. & N., Duraimurugan. (2017). Sentiment analysis on GST in social mediausing R. Journal of Advanced Research in Dynamical and Control Systems. 9. 276-282..

[12] Kumar P, Vinod Kumar K. S, P. L and S. SenthilPandi, "Enhancing Face Mask Detection Using Data Augmentation Techniques," 2023 International Conference on Recent Advances in Science and Engineering Technology (ICRASET), B G NAGARA, India, 2023, pp. 1-5, doi: 10.1109/ICRASET59632.2023.10420361

[13] A. K, J. David and K. A, "Cardiovascular Disease Prediction using Patient History and Real Time Monitoring," 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2024, pp. 1226-1233, doi: 10.1109/IDCIoT59759.2024.10467488.

[14] D. Nagendiran and S.P. Chokkalingam, "Real Time Brain Tumor Prediction Using Adaptive Neuro Fuzzy Technique," Intell. Automat. Soft Comput., vol. 33, no. 2, pp. 983-996. 2022. https://doi.org/10.32604/iasc.2022.023982