www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 13 Issue 03 March 2024, Page No. 26058-26067
ISSN: 2319-7242 DOI: 10.18535/ijecs/v13i03.4795

Predictive Analysis in Healthcare

Soumalya Ghosh Dastider*, Dr Kavitha R*

* Master of Computer Application, School of CS & IT, JAIN (Deemed-to-be) University

Abstract
This research paper explores the transformative role of predictive analysis in healthcare, with a specific focus on predicting and managing Parkinson's disease as a case study. By integrating advanced analytics and machine learning techniques, we delve into the intricate landscape of healthcare data to forecast patient outcomes, optimize resource allocation, and enhance overall healthcare decision-making. Through an exhaustive examination of literature and scientific inquiry, we illuminate the promising applications of predictive analysis in addressing the chronic and complex nature of neurodegenerative diseases like Parkinson's. The paper also discusses the challenges, ethical considerations, and future directions within the realm of predictive analysis in healthcare, offering a comprehensive perspective on its potential to revolutionize patient care and public health outcomes.

I. Introduction

Data-driven decision-making has become pivotal in the realm of contemporary healthcare, as evidenced by the increasing reliance on machine learning and predictive analytics. This shift is not merely a response to technological advancements but an acknowledgment of the transformative role these tools play in revolutionizing the healthcare sector. Predictive analytics holds paramount importance in healthcare, offering the potential to inform critical decisions and significantly improve outcomes. In addressing the complexity of the healthcare industry, stakeholders such as practitioners, academics, and policymakers are turning to data-driven insights. This paper aims to illuminate the multifaceted impact of predictive analytics in healthcare, with a special focus on patient care, resource allocation, and public health management. Drawing from an exhaustive literature analysis and scientific investigation, we explore the capabilities of predictive analytics to predict disease outbreaks, estimate patient outcomes, and optimize healthcare resource allocation [1]. As witnessed during the global response to the COVID-19 pandemic, disease outbreak prediction, a vital component of predictive analytics, has become increasingly crucial. Cutting-edge analytical tools, coupled with real-time data sources, empower healthcare practitioners to predict the spread of illnesses, implement targeted interventions, and mitigate the impact on public health [1]. Moreover, predictive analytics enhances patient care by enabling healthcare professionals to anticipate outcomes more accurately. Through the utilization of historical patient data, clinical variables, and machine learning algorithms [2], practitioners can craft more individualized treatment plans, reduce readmissions, and improve overall patient experiences. Another critical domain where predictive analytics excels is the optimal allocation of healthcare resources, a cornerstone for effective and efficient healthcare delivery. By leveraging predictive models based on historical data and real-time insights, healthcare organizations can enhance their ability to adapt to the changing demands of patient populations [3]. This study aims to comprehensively examine the relevance and broad ramifications of predictive analytics in healthcare, providing insights into its potential to reshape healthcare outcomes and decision-making. By conducting a thorough assessment of existing literature, exploring important approaches, and considering ethical considerations, we seek to shed light on the future of utilizing predictive analytics for improved healthcare outcomes and decision-making. As part of our exploration, we will use the prediction and management of Parkinson's disease as a model to exemplify the applications and challenges within the predictive analytics framework. This neurodegenerative disorder, affecting dopamine-producing neurons in the substantia nigra, serves as a poignant example of the chronic and complex nature of diseases that necessitate innovative approaches in healthcare [2]..

II. Literature Review

Predictive Analytics in Healthcare A State of the Art

The potential for predictive analytics to completely transform the healthcare sector has attracted a lot of attention in recent years. The next section of this review of the literature gives an overview of the state of predictive analytics in healthcare today and discusses pertinent research, models, and methodologies used in the areas of disease outbreak forecasting, patient outcome forecasting, and healthcare resource use.

Disease Outbreak Prediction:

Predicting disease outbreaks is a crucial part of public health. Effective resource allocation and the implementation of preventative measures can both benefit from accurate and timely projections. Numerous studies have looked at this topic; some examples include: -Google Flu Trends: Early results from Google's project to forecast flu epidemics by looking at search searches were promising [4]. Due to problems with data quality and representativeness, it struggled to maintain accuracy over time. - Machine Learning Approaches: In more recent research, illness outbreak predictions were made using machine learning methods. By utilizing climatic and epidemiological data, Fong et al. (2018) projected dengue outbreaks with great accuracy using deep learning [5].

Parkingson's Diesease Prediction:

Parkinson's disease (PD) is a neurodegenerative disorder characterized by the progressive degeneration of dopamine-producing ("dopaminergic") neurons, primarily located in the substantia nigra region of the brain. Symptoms manifest gradually over an extended period, exhibiting variability among individuals due to the diverse nature of the disease. Common symptoms include resting tremors, bradykinesia, limb rigidity, and difficulties in gait and balance.

Despite ongoing research, the etiology of Parkinson's disease remains elusive. While a definitive cure is absent, the current therapeutic landscape includes diverse options such as pharmacological interventions and surgical procedures. Although PD itself is not inherently fatal, its complications carry serious health implications. Effectively managing Parkinson's disease through predictive analysis involves gaining a nuanced understanding of the condition's progression. While a cure remains elusive, individuals with PD can achieve a satisfactory quality of life through tailored management strategies.

The need for medication in Parkinson's disease arises from depleted dopamine levels in the brain, primarily linked to the impairment of substantia nigra neurons. Importantly, symptoms manifest in the later stages of the disease when a substantial proportion of substantia nigra neurons has already suffered impairment or loss. The identification of biomarkers for PD through predictive analysis is a current research focus, aiming for early diagnosis and personalized therapeutic interventions to slow down disease progression. Present therapies predominantly address symptomatology, lacking significant impact on the underlying disease course.

In the realm of predictive analysis in healthcare, attention extends beyond motor symptoms to encompass non-motor manifestations such as apathy, depression, constipation, sleep disturbances, anosmia, and cognitive impairment. Recognizing and integrating these non-motor aspects into predictive models becomes crucial for comprehensive disease prediction and management.

The early identification of Parkinson's disease signs is imperative for timely intervention within the predictive analysis framework. Ten predictive signs include resting tremors, micrographia, anosmia, sleep disturbances, impaired movement, constipation, a soft or low voice, facial masking, dizziness, and stooping. While a single sign may not definitively predict PD, the concurrence of multiple signs prompts proactive medical assessment within the predictive analytics context.

Individuals diagnosed with Parkinson's disease are encouraged to engage in collaborative predictive healthcare analysis with professionals to formulate a comprehensive plan for optimal health maintenance. This may involve consultations with predictive analysts, neurologists, occupational therapists, physical therapists, and speech therapists. Initiating a predictive exercise program is recommended to preemptively address symptom progression. The support of family and friends plays a pivotal role in navigating the challenges posed by Parkinson's disease within the predictive healthcare paradigm.

III. Ethical Consideration

It is important to note that as predictive analytics are increasingly used in healthcare, ethical issues such patient privacy, data security, and algorithmic bias have surfaced [8]. Predictive model responsibility is a subject of continuing study and discussion. A viable option for improving disease outbreak prediction, patient outcome forecasting, and healthcare resource use is provided by predictive analytics in healthcare. These developments are made possible by cutting-edge machine learning and deep learning techniques, which are frequently used on a vast amount of healthcare data. However, there are certain difficulties in the subject, notably with regard to ethical and privacy issues. To overcome these problems and expand the application of predictive analytics in healthcare, more research and multidisciplinary cooperation are required.

Patient privacy and data security, as well as the possibility of bias in predictive models, are just a few of the ethical issues that are emerging as healthcare predictive analytics continues to develop and pervade the healthcare scene. To ensure the ethical and effective application of predictive analytics in healthcare, it is essential to address these ethical issues. Patient privacy is one of the most important ethical issues with healthcare predictive analytics. Sensitive, personally identifiable information is frequently included in the enormous quantity of data used in prediction models. It can be difficult to strike a balance between using this data to improve medical care and preserving people's privacy. Protecting patient privacy requires careful de-identification procedures as well as strong data protection safeguards including encryption and access controls [19]. Data protection: Another urgent ethical concern is the protection of healthcare data. Patients may suffer serious injury as a result of breaches or illegal access to predictive analytics systems, and their sensitive health information may even be made available to criminals. To prevent data breaches and preserve public trust, it is essential to implement strong cybersecurity safeguards and adhere to data protection laws like HIPAA (Health Insurance Portability and Accountability Act) [20]. Bias in prediction Models: Due to the potential for perpetuating healthcare inequities and inequality, bias in prediction models raises serious ethical questions. As a result, some patient groups may receive biased forecasts or recommendations from models that were trained on historical data. Fairness assessments, in-depth model evaluations, and actions like data reweighting or algorithm changes are all part of the process of addressing bias [8]. Albeit complicated, ethical issues relating to healthcare predictive analytics are crucial. Utilizing the advantages of predictive analytics while preserving the ethical values of patient autonomy, beneficence, and justice requires safeguarding patient privacy, guaranteeing data security, and eliminating bias in prediction models. Healthcare practitioners, academics, and politicians in the sector are committed to addressing these ethical issues on an ongoing basis.

IV. Methodology

Data Collection:

The caliber and range of data sources are crucial in healthcare predictive analytics for producing precise and valuable insights. The predictive potential of models is increased by utilizing many data sources. These sources include electronic health records (EHRs), patient demographics, environmental information, and historical epidemic data. Electronic health records (EHRs): EHRs include thorough information on medical histories, treatments, prescriptions, and test findings. They are a priceless source of patient-specific data. Because they offer longitudinal data that enables the creation of prediction models for identifying at-risk patients, EHRs are crucial for patient outcome forecasting [2]. EHRs also make it easier to extract crucial clinical characteristics that may be used to forecast outcomes, readmissions, and disease progression. Patient Demographics: Predictive analytics models depend on data from patient demographics, including age, gender, socioeconomic status, and geography. These data points may be used to identify vulnerable populations, comprehend gaps in healthcare access and outcomes, and design interventions accordingly [9]. For instance, financial position may have an impact on how healthcare resources are used, while age and gender are important factors in predicting the incidence of some diseases. Environmental data, which includes elements like air quality, temperature, and humidity, is being included into healthcare predictive analytics models, particularly for the prediction of disease outbreaks. The transmission of infectious illnesses is significantly influenced by environmental variables. Real-time environmental data gathered from a variety of sources, including weather stations and satellites, has been shown to be useful in forecasting disease outbreaks like dengue fever and influenza [5]. Historical Outbreak Data: Historical outbreak data are a gold mine of knowledge for predicting disease outbreaks. These databases contain information on previous pandemics, epidemics, and outbreaks, providing insights into the dynamics and patterns of infectious illnesses. Researchers and public health professionals can develop models that take into account previous trends by examining historical data, which can result in more precise projections and timely interventions [10]. Predictive analytics' performance in the healthcare industry depends critically on the accessibility and integration of many data sources. The information landscape is enriched by these sources, which also include electronic health records, patient demographics, environmental data, and historical outbreak data. This allows for more thorough and precise predictions of disease outbreaks, patient outcomes, and the use of healthcare resources

In Parkingson’s Disease prediction, the dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver,Colorado, who recorded the speech signals.

The original study published the feature extraction methods for general voice disorders.

There are a total of 195 instances in the life domain, with 23 characteristics recorded on June 26, 2008. No related assignments regarding missing values in the classification were identified. The file size is 39.7 KB.

Methodology:

Prediction accuracy and dependability in healthcare predictive analytics are highly dependent on the approach used. A variety of machine learning and predictive analytics approaches may be used to handle the complex objectives of forecasting disease outbreaks, patient outcomes, and resource usage in the healthcare industry. Regression analysis, time series forecasting, classification algorithms, and deep learning are some of the options. Regression analysis is a frequently used method in the context of predicting patient outcomes. For instance, linear regression may be used to forecast continuous variables like the risk of patient death. Healthcare practitioners may estimate patient outcomes with a high degree of precision by identifying pertinent aspects, such as clinical factors and demographics, and creating regression models [11]. Time series data are frequently used in disease outbreak prediction, and past data points are important for predicting upcoming occurrences. In order to model temporal patterns and trends in epidemic data, time series forecasting techniques like ARIMA (AutoRegressive Integrated Moving Average) and exponential smoothing are used. When taking into account the dynamics of infectious illnesses and their seasonality, these strategies are very helpful [12]. Classification algorithms are essential for identifying individuals who are at risk of readmission or who have certain diseases. Binary classification issues are frequently solved using support vector machines, decision trees, and logistic regression. These algorithms can profit from methods like feature engineering since they need feature selection to find the most pertinent predictors [13]. Deep learning is a subset of machine learning that has become very popular in predictive analytics due to its capacity to evaluate complicated, unstructured data, such as textual entries in electronic health records or medical pictures. For applications like image-based illness detection or sentiment analysis of medical records, deep neural networks, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), are employed to analyze data sequences or pictures, respectively [14]. The technique must include preprocessing processes, feature selection, and data cleaning: - Data normalization, addressing missing values, and handling outliers are all part of preprocessing. Missing values are frequent in healthcare data and must be handled carefully to maintain data quality. You can use imputation techniques like mean, median, or regression imputation. When combining data from many sources with differing scales, data normalization, such as standardization, is crucial. - <u>Feature Selection</u>: Feature selection is the process of choosing the variables that will provide prediction models the most useful information. The performance of the model may be determined by using methods such as feature significance analysis, recursive feature removal, and correlation analysis. - <u>Data Cleaning:</u> Data cleaning includes the elimination of duplicate records, the repair of inaccurate entries, and the handling of inconsistent data entry. Data cleaning is an essential step to guarantee data integrity since anomalies and outliers can have a substantial impact on the performance of predictive models. Depending on the precise healthcare predicting job at hand, one may choose from a variety of predictive analytics methodologies, including regression analysis, time series forecasting, classification algorithms, and deep learning. To make sure that data are dependable and pertinent for developing precise models, preprocessing, feature selection, and data cleaning are equally important.

The predictive analysis methodology for Parkinson's disease using the Python script involves several key steps. First, data preprocessing is performed, including handling missing values, exploring data distributions, and identifying outliers using distribution plots and boxplots. The dataset is then oversampled using the RandomOverSampler to address class imbalance.

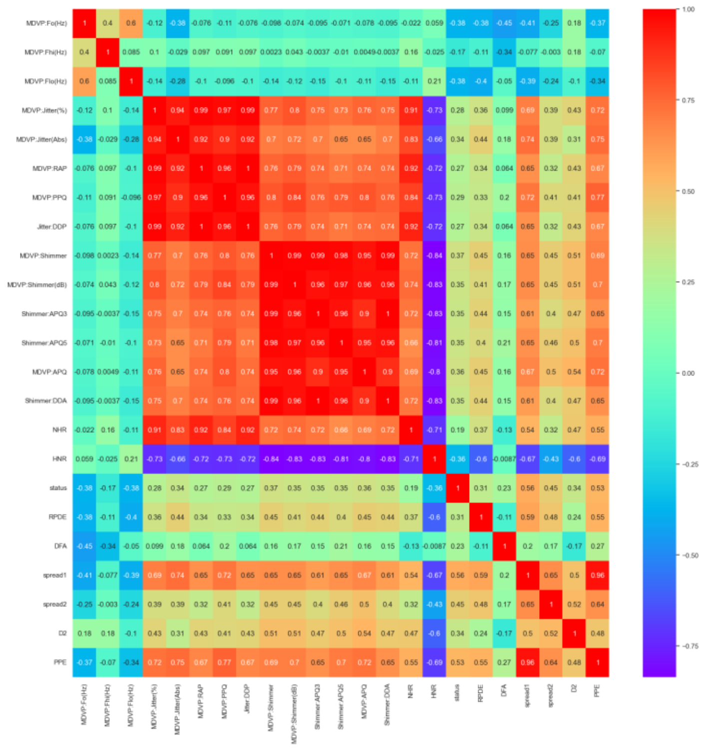
	name	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	MDVP:Shimmer(dB)	Shimm
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.000070	0.00370	0.00554	0.01109	0.04374	0.426	
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.000080	0.00465	0.00696	0.01394	0.06134	0.626	
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.000090	0.00544	0.00781	0.01633	0.05233	0.482	
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.000090	0.00502	0.00698	0.01505	0.05492	0.517	
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.000110	0.00655	0.00908	0.01966	0.06425	0.584	
5	phon_R01_S01_6	120.552	131.162	113.787	0.00968	0.000080	0.00463	0.00750	0.01388	0.04701	0.456	
6	phon_R01_S02_1	120.267	137.244	114.820	0.00333	0.000030	0.00155	0.00202	0.00466	0.01608	0.140	
7	phon_R01_S02_2	107.332	113.840	104.315	0.00290	0.000030	0.00144	0.00182	0.00431	0.01567	0.134	
8	phon_R01_S02_3	95.730	132.068	91.754	0.00551	0.000060	0.00293	0.00332	0.00880	0.02093	0.191	
9	phon_R01_S02_4	95.056	120.103	91.226	0.00532	0.000060	0.00268	0.00332	0.00803	0.02838	0.255	
10	phon_R01_S02_5	88.333	112.240	84.072	0.00505	0.000060	0.00254	0.00330	0.00763	0.02143	0.197	

Informations about attributes in Dataset:

The matrix is comprised of columns with distinct attributes:

- name: Subject identifier in ASCII and recording number
- MDVP:Fo(Hz): Average vocal fundamental frequency
- MDVP:Fhi(Hz):Maximumvocal fundamental frequency

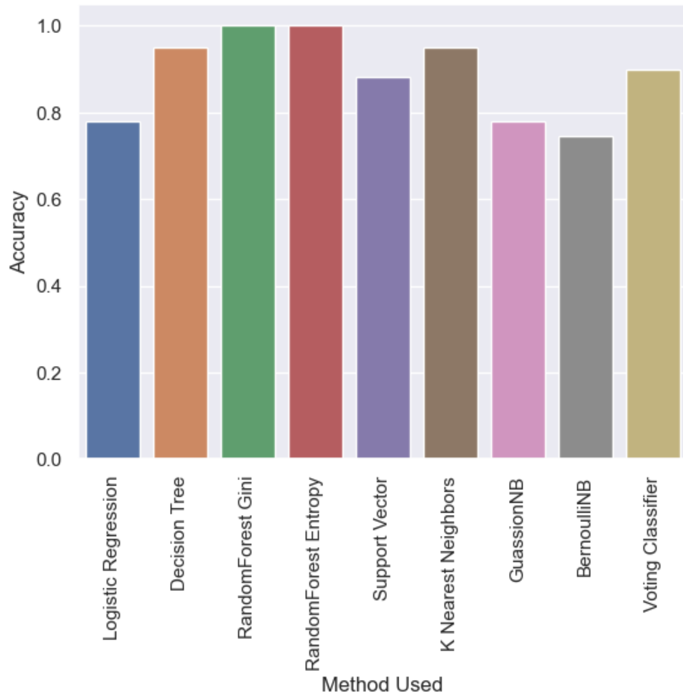
- MDVP:Flo(Hz):Minimum vocal fundamental frequency
- MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:RAP, MDVP:PPQ, Jitter:DDP: Various metrics indicating the variation in fundamental frequency
- MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ, Shimmer:APQ5, MDVP:APQ,Shimmer:DDA: Various metrics reflecting the variation in amplitude
- NHR, HNR: Two measures denoting the ratio of noise to tonal components in the voice
- status: Health condition of the subject (1 for Parkinson's, 0 for healthy)
- RPDE, D2: Two nonlinear dynamical complexity measures
- DFA: Signal fractal scaling exponent
- spread1, spread2, PPE: Three nonlinear measures of fundamental frequency variation



Feature scaling is applied using MinMaxScaler, and Principal Component Analysis (PCA) is employed to reduce dimensionality while retaining 95% of the variance. Subsequently, the dataset is split into training and testing sets.

Multiple machine learning models are trained and evaluated, including Logistic Regression, Decision Tree, Random Forest with Gini and Entropy criteria, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), and Bernoulli Naive Bayes (BNB). The models' performances are assessed using accuracy scores, and a Voting Classifier is implemented to combine their predictions.

```
             Method Used   Accuracy
0    Logistic Regression   0.779661
1          Decision Tree   0.949153
2      RandomForest Gini   1.000000
3   RandomForest Entropy   1.000000
4         Support Vector   0.881356
5    K Nearest Neighbors   0.949153
6             GuassionNB   0.779661
7            BernoulliNB   0.745763
8      Voting Classifier   0.898305
Axes(0.125,0.11;0.775x0.77)
```



The evaluation includes confusion matrices and classification reports for each model on both the training and testing sets. Additionally, a Receiver Operating Characteristic (ROC) curve is plotted for the Random Forest model with Entropy criterion.

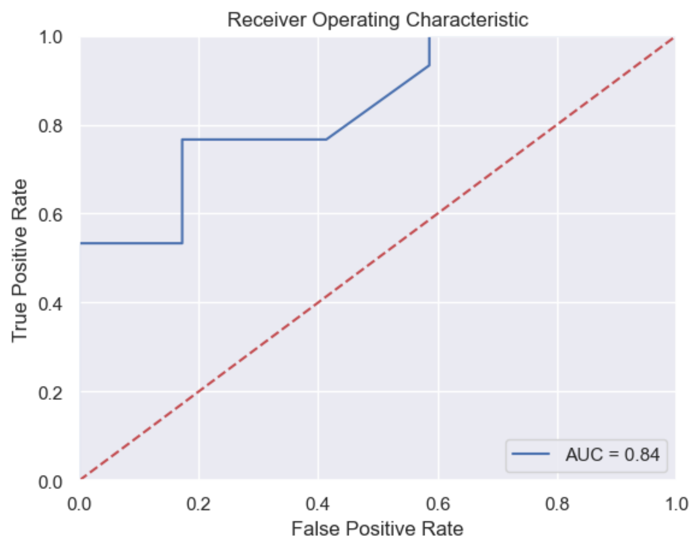
Confusion Matrix for KNN:

```
[[118   0]
 [ 12 105]]
*************************
[[29  0]
 [ 3 27]]
```

Classification Report for KNN:

```
              precision    recall  f1-score   support

           0       0.91      1.00      0.95       118
           1       1.00      0.90      0.95       117

    accuracy                           0.95       235
   macro avg       0.95      0.95      0.95       235
weighted avg       0.95      0.95      0.95       235

*************************************************************************
              precision    recall  f1-score   support

           0       0.91      1.00      0.95        29
           1       1.00      0.90      0.95        30

    accuracy                           0.95        59
   macro avg       0.95      0.95      0.95        59
weighted avg       0.95      0.95      0.95        59
```

The final step involves cross-validation using the cross_val_score function. Overall, the methodology encompasses data preprocessing, oversampling, feature scaling, dimensionality reduction, model training, evaluation, and cross-validation to ensure robust predictive analysis for Parkinson's disease classification.



V. Challenges and limitations

Although predictive analytics has a lot of potential for the healthcare industry, it is not without its difficulties and restrictions. The correctness and completeness of the input data have a significant influence on the dependability of the model, making data quality a major problem. Electronic health records that are inconsistent or lacking information might result in inaccurate forecasts and choices. Additionally, it is frequently difficult to understand models, especially when using sophisticated machine learning techniques like deep neural networks. Forging trust and making wise selections, it's essential to comprehend a model's basic logic. Although interdisciplinary cooperation is crucial, it might be difficult to promote since healthcare predictive analytics call for knowledge in medicine, data science, and ethics. Collaboration across these disparate areas continues to be very difficult to foster. To guarantee ethical and fair use of predictive analytics in healthcare, continual attention is required to issues like algorithmic bias and data privacy [21][8][22].

VI. Future

The integration of artificial intelligence (AI) in numerous healthcare domains is one of the interesting discoveries that will shape the future of healthcare predictive analytics. AI-driven algorithms are assisting radiologists and pathologists in medical imaging by assisting in the early identification of illnesses like cancer, increasing the accuracy of diagnoses and the quality of patient outcomes. With AI contributing in the identification of genetic markers linked to illness risk and treatment response, genomic data is another area for predictive analytics. Additionally, improvements in deep learning and natural language processing may make it possible to draw important conclusions from unstructured clinical data, including notes from doctors and patient records. Predictive analytics is set to play an increasingly important role in healthcare as these trends develop, influencing how illnesses are identified, treated, and managed [14][23].

VII. Conclusion and Recommendations

Predictive analytics in healthcare is a potent tool that may enhance patient outcomes, optimize resource allocation, and support disease management. Its potential benefits are significant despite issues with data quality, model interpretability, and ethical considerations. Healthcare practitioners, legislators, and academics must collaborate to solve these issues in order to improve this profession. It is advised that healthcare organizations make investments in data governance and infrastructure to guarantee the integrity and security of patient data. To improve the multidisciplinary abilities required for the effective application of predictive analytics, training programs should be created. In addition, continued research is required to improve models, reduce bias, and investigate novel applications in genomics and AI. These suggestions can help healthcare predictive analytics reach its full potential in modernizing decision-making and healthcare delivery [9].

Acknowledgement

I would like to express our sincere gratitude to all those who have contributed to the successful completion of this research project. Their support, guidance, and encouragement have been invaluable.

First and foremost, I extend my deepest appreciation to my mentor, Dr Kavitha R, for her unwavering support, insightful guidance, and patience throughout the research process. Her expertise and dedication played a critical role in shaping the direction of this study.

My sincere thanks go to Jain (Deemed-to-be) University for providing access to resources and facilities that facilitated our data collection and analysis.

Last but not least, we are thankful to our families and friends for their unwavering encouragement and understanding during this demanding research journey.

This research would not have been possible without the collective efforts and support of those mentioned above. I acknowledge their contributions with sincere gratitude.

References

1. Cho, S., Sohn, C. H., Jo, M. W., Shin, S. Y., & Lee, J. H. (2013). Correlation between national influenza surveillance data and google trends in South Korea. PloS One, 8(12), e81422.
2. Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Dean, J. (2018). Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine, 1(1), 18.
3. Levin, S., Jain, S., & Kruglov, D. (2016). Predictive analytics in healthcare: A roadmap for the health sector. McKinsey & Company.
4. Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. Nature, 457(7232), 1012-1014.
5. Fong, S., Nguyen, H. V., Lee, C., Deivasigamani, R., Goh, J., Yee, W., & Soewito, B. (2018). Deep learning for dengue outbreak prediction. Computers in Biology and Medicine, 104, 24-31.
6. Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., ... & An, M. (2011). Risk prediction models for hospital readmission: A systematic review. JAMA, 306(15), 1688-1698.
7. Green, L., Kim, E. H., Lee, S., & Khuri, J. Y. (2018). Predicting emergency department wait times in real-time. Health Care Management Science, 21(3), 347-358.
8. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.
9. Dhar, V. (2013). Data science and prediction. Communications of the ACM, 56(12), 64-73.
10. Lippi, G., Henry, B. M., & Sanchis-Gomar, F. (2020). COVID-19: Unravelling the data. International Journal of Environmental Research and Public Health, 17(10), 3648.
11. Goldberger, Z. D., Rong, J., & Winkler, S. (2018). Developing Electronic Health Record Algorithms That Accurately Identify Patients With Implantable Cardiac Devices in a Large Integrated Healthcare System. Journal of the American Heart Association, 7(17), e008137.
12. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). Time Series Analysis: Forecasting and Control. John Wiley & Sons.

13. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression. John Wiley & Sons.
14. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.
15. World Health Organization (2020). Coronavirus disease (COVID-19) technical guidance: Surveillance and case definitions. Retrieved from https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON262
16. Hethcote, H. W. (2000). The mathematics of infectious diseases. SIAM Review, 42(4), 599-653.
17. Block, P., Hoffman, M., Raabe, I. J., Dowd, J. B., Rahal, C., Kashyap, R., & Mills, M. C. (2020). Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. Nature Human Behaviour, 4(6), 588-596.
18. Shaman, J., Karspeck, A., Yang, W., Tamerius, J., & Lipsitch, M. (2013). Real-time influenza forecasts during the 2012-2013 season. Nature Communications, 4(1), 2837.
19. Raisaro, J. L., Troncoso-Pastoriza, J. R., Misbach, M., Quer, G., Hedenius, M., Elger, B. S., ... & Troncoso, C. (2018). PEPs: Privacy-preserving epidemic prediction in connected populations. PLOS Computational Biology, 14(3), e1005993.
20. McConnell, J. W., & Data Privacy Group. (2013). Health data privacy: The promise and the peril. Health Affairs, 32(11), 1979-1985.
21. Car, J., Sheikh, A., Wicks, P., Williams, M. S. (2015). Beyond the hype of big data and artificial intelligence: building foundations for knowledge and wisdom. BMC Medicine, 13, 260.
22. Wang, F., Casalino, L. P., Khullar, D. (2018). Deep Learning in Medicine—Promise, Progress, and Challenges. JAMA Internal Medicine, 178(3), 293-294.
23. Vassy, J. L., Christensen, K. D., Schonman, E. F., Blout, C. L., Robinson, J. O., Krier, J., ... & Rehm, H. L. (2019). The Impact of Whole-Genome Sequencing on the Primary Care and Outcomes of Healthy Adult Patients: A Pilot Randomized Trial. Annals of Internal Medicine, 170(6), 152-159.