OPEN ACCESS JOURNALS
Indiana Journal of Multidisciplinary Research
Abbriviate Tittle- Ind J Multi Res
ISSN *(Online)*- 2583-3820
Journal Homepage Link- https://indianapublications.com/Journals/IJMR
DOI: https://doi.org/10.5281/zenodo.12671529

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Article

Volume-04|Issue-03|2024

Big Data Semantic Management for Disease Prediction

Kuldeep Jain[1], Aadithya V Jamadagni[2], Adarsh S[3], Bharath K[4], Deepak N A[*5]

[1,2,3,4] Student, Computer Science and Engineering, RVITM, Bengaluru, Karnataka, India
[5]Associate Professor, Computer Science and Engineering, RVITM, Bengaluru, Karnataka, India

Article History
Received: 20.05.2024
Accepted: 05.06.2024
Published: 30.06.2024

Citation
Kuldeep Jain, Aadithya V Jamadagni, Adarsh, S., Bharath, K., Deepak, N. A. (2024). Big Data Semantic Management for Disease Prediction. *Indiana Journal of Multidisciplinary Research*, *4*(3), 87-90.

Abstract: Traditional diagnostic techniques for health conditions like Diabetes, Alzheimer's, Skin Cancer swells, Lung including Cancer, and also Malaria typically require invasive procedures with biomarker analysis. These methods usually take time, are expensive, and tend to have some limits in terms of accuracy. Nonetheless, recent progress in machine learning, for the analysis of clinical data, and Convolutional Neural Networks (CNNs) for processing medical images have opened up possibilities for more complex diagnostic models. There is a rising need for an approach able to correctly diagnose a variety of diseases using different data modalities. Our model suggests exclusive machine learning algorithms for each disease, for instance, the Random Forest Classifier intended for Diabetes. This involves focusing on clinical data for precise predictions. Convolutional Neurological Networks (CNNs) are utilized for Alzheimer's and Skin Cancer, utilizing medical images, alongside for Malaria, the scrutiny of blood smear images. The model has a menu-driven interface for the easy input of data (images, reports) and it provides clear diagnostic outputs, bolstered by confidence scores intended to improve the decision- making process in clinical environments.

The model has been tried out using multiple medical datasets, displaying high accuracy in the diagnosis and classification of illnesses, especially within the specified diseases. It delivers efficient processing times and a user-friendly interface that simplifies integration into clinical workflows. The potential for real-world applications of this model is considerable, holding the promise of assisting healthcare professionals in providing accurate and prompt diagnoses, thus enhancing patient care. Subsequent publications will delve into the specific functionalities and technical aspects of the model.

Keywords: Convolution Neural Network (CNN), Random Forest Classifier, Diabetes, Alzheimer, Skin Cancer, Lung Cancer, Malaria.

INTRODUCTION

In the upcoming endeavor, our focus lies in pioneering advancements within the medical domain and enhancing diagnostic procedures. By harnessing the power of various AI/ML techniques, our primary objective entails streamlining patient diagnoses. Through these cutting-edge technologies, we strive to make accurate predictions relating to an array of diseases and cancers among a specified group of individuals. This groundbreaking initiative not only directly impacts the healthcare sector but also significantly benefits society as a whole. The incorporation of an innovative diagnostic framework that combines machine learning and data science represents a pivotal step towards tackling prevalent illnesses. Not only does this facilitate improved patient care through prompt and precise prognoses, but it also fosters a culture of technological progress. By placing a strong emphasis on ethical considerations, our project ensures alignment with societal norms, thereby nurturing trust among stakeholders. Collaborative endeavors with esteemed medical professionals provide invaluable opportunities for multidisciplinary research, unveiling potential breakthroughs in cost-effective healthcare solutions. These collaborative efforts underscore the project's profound societal influence and significance. The innovative approach adopted in this project marks a significant milestone in the realm of healthcare diagnostics, with far-reaching implications for diverse industries, communities, and institutions alike.

MATERIALS AND METHODS

The methodology for the Disease Prediction project can be summarized as follows:

Data Gathering: Data is being collected from a well-known platform called Kaggle.com. This is a popular site for accessing and obtaining various datasets. The collected data is specifically focused on diseases such as Diabetes, Alzheimer's, Skin cancer lesion, Malaria, and Lung cancer.

Data Cleaning: Before moving forward with the analysis, the collected data needs to go through a pre-processing stage to make sure it's of top-notch quality and suitable for training the machine learning models. Key steps in this process involve taking care of missing values, getting rid of duplicates, and conducting data normalization and feature scaling.

Model Choice: The project chooses from a range of pre trained models that utilize different machine learning algorithms dedicated to disease prediction tasks. The Random Forest Classifier is chosen for tasks where ensemble methods are at their best while Convolutional Neural Networks (CNNs) are applied for image-based

***Corresponding Author:** Deepak N A

diagnostics. The utilization of TensorFlow along with Keras enables the rapid deployment of deep learning models. The extensive library available within TensorFlow enhances complex predictions.

Training and Validation: The pre-processed data is then undergoes a division into training and testing sets. Models are trained using the training dataset, and their performance evaluation is conducted using the testing dataset. Accuracy serves as the main evaluation metric to assess each model's performance.

Model Implementation: Seamless deployment of machine learning models can be achieved using Python and Flask. The process involves setting up a lightweight API for model integration. Flask acts as the interface for incorporating Python-based models, thereby enabling accessible and efficient real- time predictions.

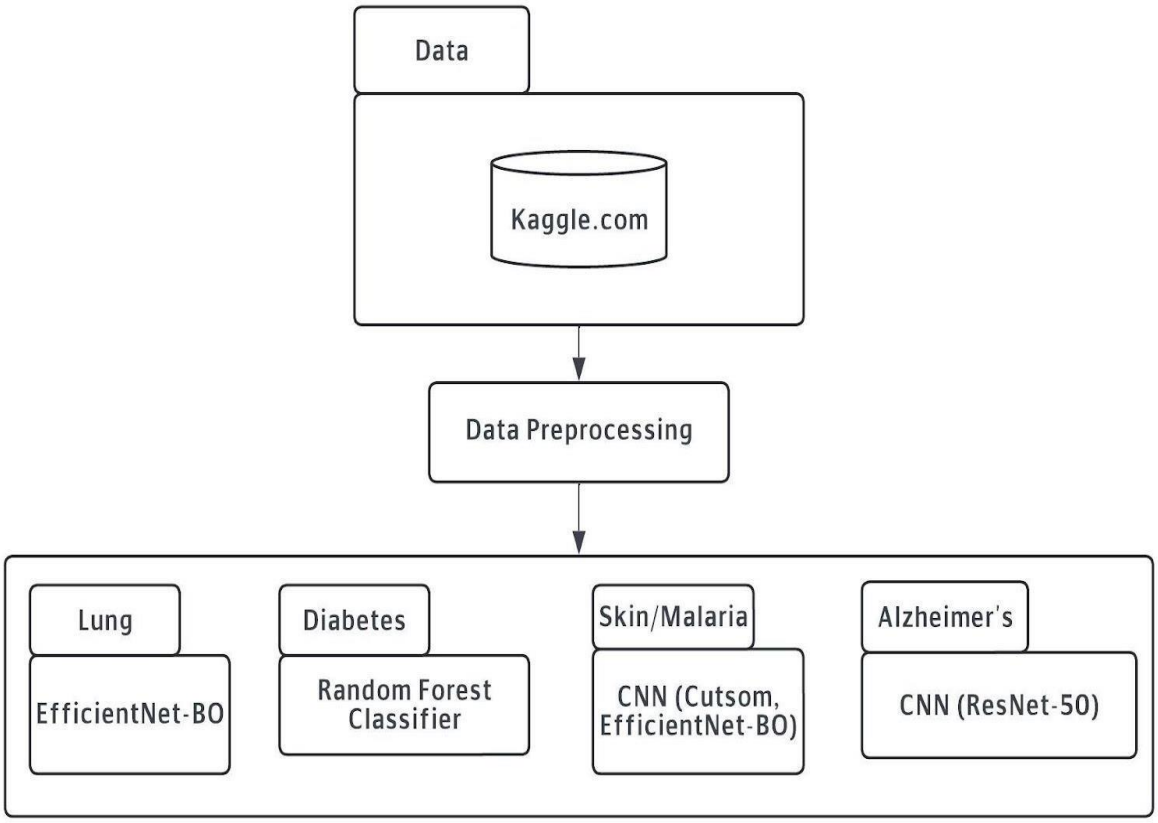

Figure 1: Medical Diagnosis Model Workflow

PROPOSED SYSTEM

In an existing system, the models are not implemented with TensorFlow and Keras, but in the proposed system there are multiple diseases added to the existing system that is implemented by neural networks with the help of TensorFlow and Keras. We use the latest techniques like data standardization to standardize the data, Label encoding techniques to transform text data into numerical data, and dimensionality reduction to reduce the features with loss of information from the data. We use the algorithms that are perfectly suitable for the dataset, we take simple models to increase the model performance.

The proposed system is a comprehensive disease prediction project that utilizes machine learning algorithms, including Random Forest classifier, Convolution Neural network (CNN), TensorFlow, and Keras to predict multiple diseases such as diabetes, Alzheimer's, Skin Cancer lesion, Malaria, and Lung cancer. The system aims to provide accurate disease predictions based on input parameters and a user-friendly interface developed using Flask and Python. Data for the models is collected from the Kaggle platform, a popular data science community, and is preprocessed to ensure its quality and suitability for training the models. The preprocessed data is then used to train the respective machine learning algorithms specific to each disease. The pre-trained models are tested to evaluate their accuracy in disease prediction

The system employs the Random Forest classifier algorithm to predict diabetes, achieving an accuracy of 95.4%. This indicates that the Random Forest classifier model can accurately identify the presence or absence of diabetes in patients, aiding in early detection and effective management!

For Alzheimer's disease prediction, the system uses the CNN algorithm (ResNet 50V2, VGG19) with the accuracy for each architecture at 89.5% and 83.2%. This high accuracy demonstrates the capability of the

***Corresponding Author:** Deepak N A

CNN model to distinguish individuals with Alzheimer's disease from healthy individuals. Malaria disease prediction is performed using the CNN algorithm the architecture used in this is EfficientNet B0 which gives an accuracy of 98%. This demonstrates the efficient and robust predictiveness of the model

For Skin cancer lesions, the algorithm used is CNN the architecture behind the pre-train model is DenseNet which has proven effective in skin cancer detection by analyzing dermatoscopic images. Leveraging its layers, DenseNet can identify subtle patterns in skin lesions that are indicative of various cancer types! This significantly enhances diagnostic accuracy of 87%, aiding clinicians in early and accurate skin cancer identification

The proposed system employs a EfficientNet B0 model for precise segmentation and prediction of lung cancer from CT scans, achieving remarkable accuracy in identifying malignant tissues. The accuracy of 81% enhances the reliability of diagnoses and supports effective treatment planning

INPUT AND OUTPUT DESIGN

Input: The Disease prediction system necessitates user input in the form of a specific parameter; the website allows the users to input their data (input parameter) then the backend connects with a pre train model which calls the actual function for which the data is collected from the user. It should ensure that the parameters requested are relevance for disease prediction and then it moves into the output phase where the precise prediction is displayed.

Output: The result of the prediction is displayed on the page and it displays what kind of disease it has predicted with the given input data. The user has a better interactive interface which is much more streamlined to the result. confusion, the user can easily understand the output and make informed decisions based on the information provided.

SYSTEM DESIGN

The architecture diagram for the disease prediction web application.

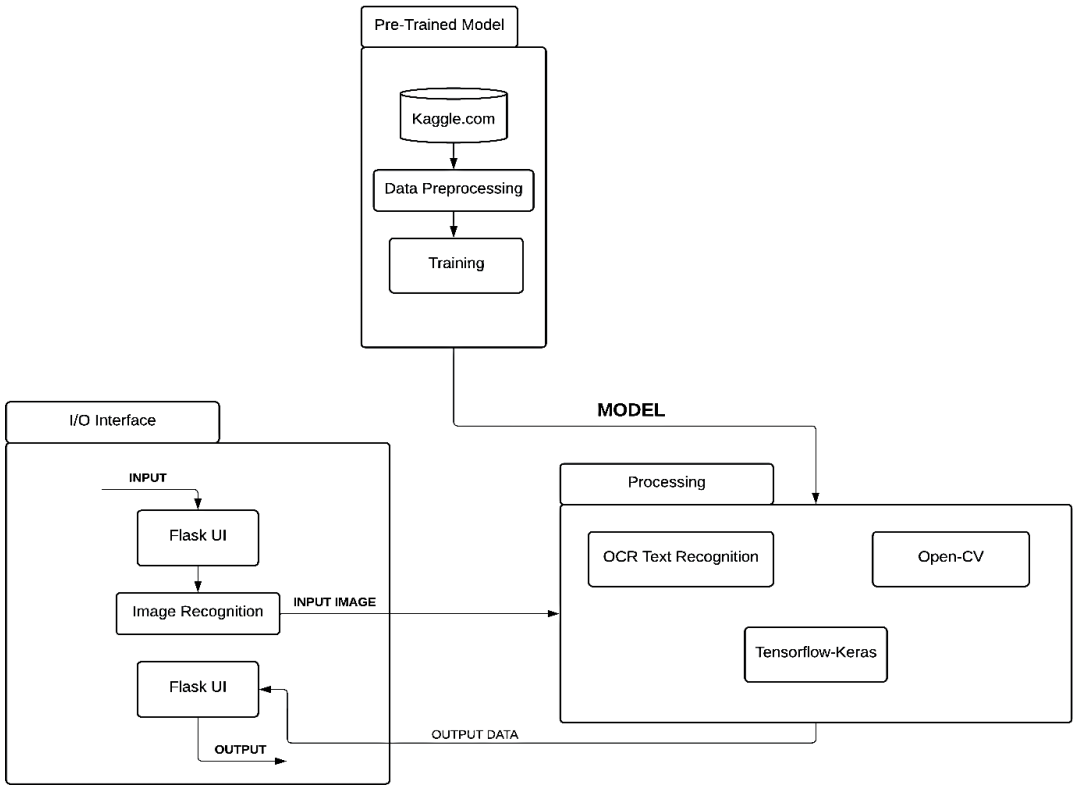

Figure 2: Medical Diagnosis System Architecture

RESULTS

This initiative tremendously transforms user interactions with healthcare services by providing a highly user-friendly interface that simplifies the process of uploading medical documents. This, in turn, leads to the acceleration of accurate diagnoses for various diseases. Through the utilization of cutting-edge technologies including Random Forest Classifiers and Convolutional Neural Networks (CNNs), along with contour detection techniques, the system ensures exceptional precision in the identification of diseases. Moreover, the project meticulously adheres to strict ethical standards, placing great emphasis on robust data privacy and security for all users. Together, these advancements in technology and ethical standards reshape the landscape of healthcare accessibility for

***Corresponding Author:** Deepak N A

users worldwide! We're excited to witness the positive impact it will have on the healthcare industry.

Sl.No	Disease Name	Algorithm Name	Accuracy
1	Diabetes	Random Forest Classifier	94.5%
2	Alzheimer's	Convolutional Neural Networks	96%
3	Skin Cancer	Convolutional Neural network	96%
4	Lung Cancer	EfficientNet B0	90%
5	Malaria	EfficientNet B0	98%

CONCLUSION

The development of a web application as chatted in this report is aimed for addressing the predictive analysis of diseases such as skin cancer, breast cancer, diabetes, Alzheimer's, and malaria. The application is designed to integrate and apply advanced computational models and also machine learning techniques in providing real time predictive insights for these diseases. The project's success will be measured by it is ability to process diverse data inputs accurately and efficiently, providing valuable support in the medical field. This endeavor represents a significant stride in merging theoretical computation methods with practical applications, potentially revolutionizing the approach to disease prediction and healthcare.

Furthermore, this project is expectably to contribute significantly to educational purposes, enhancing the understanding of disease dynamics and predictive modeling among users. Ultimately, the success of this application will be a testament to the effectively application of theoretical computational concepts in practical, health related scenarios, bridging the gap between data-driven research and user- centric health solutions!

REFERENCES

1. Patil, V., Madgi, M., & Kiran, A. (2022). Early prediction of Alzheimer's disease using convolutional neural network: a review. *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, *58*(1), 2 8 https://doi.org/10.1186/s41983 022 00571 w
2. Rahman, A., Zunair, H., Rahman, M., Yuki, J. Q., Biswas, S., Alam, M. A., Alam, N. B., & Mahdy, M. R. C. (2019). Improving Malaria Parasite Detection from Red Blood Cell using Deep Convolutional Neural Networks. *arXiv (Cornell University)*,5 28 https://arxiv.org/pdf/1907.10418
3. Femil, J. J., & Jaya, T. (2023). An efficient hybrid optimization for skin cancer detection using the PNN classifier. *Computer Systems Science and Engineering*, *45*(3), 3-12. https://doi.org/10.32604/csse.2023.032935
4. Xue, T., Jiang, M., Liu, Y., Hu, Q., Zhu, B., Hu, J., Guo, W., Wu, X., Xiong, Y., Shi, X., Zhang, X., Han, X., Li, W., Tong, R., & Long, E. F. (2023). Predicting three month fasting blood glucose and glycated hemoglobin changes in patients with type 2 diabetes mellitus based on multiple machine learning algorithms. *Scientific Reports*, *13*(1), 3-11. https://doi.org/10.1038/s41598023 43240 5
5. K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer," *Neuroscience Informatics*, 2(4), 2 8. https://www.sciencedirect.com/science/article/pii/S 2772528621000340
6. MULTIPLE DISEASE PREDICTION USING MACHINE LEARNING, DEEP LEARNING
7. AND STREAM LIT. (2023). *International Research Journal of Modernization in Engineering Technology and Science*. https://doi.org/10.56726/irjmets42818
8. Mohit, I., Kumar, K. S., Reddy, U. a. K., & Kumar, B. S. (2021). An Approach to detect multiple diseases using machine learning algorithm. *Journal of Physics: Conference Series, 2089(1)*, 012009.https://doi.org/10.1088/17426596/2089/1/ 012009

***Corresponding Author:** Deepak N A