Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 5, May 2024

Disease Detection System Using Machine Learning: A Survey

Aryan Pathak
Department of Computer Engineering
Dr. D. Y. Patil Institute of Technology
Pune, Maharashtra, India

Abhijeet Kokare
Department of Computer Engineering
Dr. D. Y. Patil Institute of Technology
Pune, Maharashtra, India

Aryesh Nair
Department of Computer Engineering
Dr. D. Y. Patil Institute of Technology
Pune, Maharashtra, India

Prof. Jithina Jose
Department of Computer Engineering
Dr. D. Y. Patil Institute of Technology
Pune, Maharashtra, India

***Abstract*—This survey underscores the critical role of early and accurate disease diagnosis in improving patient outcomes. It explores the innovative integration of Convolutional Neural Networks (CNNs), a powerful deep learning technique, to enhance disease detection accuracy for conditions like liver disorders, hepatitis, heart disease, diabetes, and chronic kidney disease. Leveraging datasets from Kaggle, the survey proposes a Disease Detection System (DDS) equipped with a Graphical User Interface (GUI) to empower medical professionals. The survey analyzes the four main causes of diseases: infection, deficiency, heredity, and organ dysfunction. The proposed DDS utilizes CNNs, which excel at extracting intricate patterns from image data. Unlike traditional machine learning algorithms, CNNs automatically learn these patterns from raw medical images (e.g., X-rays, CT scans) without the need for manual feature extraction. This inherent ability to learn from vast amounts of data allows CNNs to achieve remarkable accuracy in disease detection. The Python-developed GUI offers a user-friendly tool for doctors to efficiently screen and diagnose patients. Building on this success, the survey explores the broader landscape of disease detection encompassing various medical data types, including images and bio-signals. It analyzes limitations of current methods and emphasizes the potential of deep learning and advanced neural networks to address them. By critically evaluating past approaches, the survey paves the way for future research in disease diagnosis, ultimately aiming to improve patient outcomes.**

***Keywords*—Convolutional Neural Networks (CNNs), Deep learning, Machine learning algorithms, Feature extraction, Infection.**

INTRODUCTION (HEADING 1)

In regions with limited medical staff, diagnosing diseases can be a challenge. Unclear symptoms and expensive tests can make accurate diagnosis difficult. Traditional methods might also lead to misdiagnosis due to a lack of clear indicators or the presence of rare diseases. This can result in unnecessary treatments, impacting both patient health and the healthcare system's resources.

Convolutional Neural Networks (CNNs) offer a powerful tool to address these challenges. Unlike traditional methods that rely on pre-defined features, CNNs excel at automatically identifying crucial patterns within medical data, such as X-rays or scans. This eliminates the time-consuming and potentially subjective process of manual feature extraction. Furthermore, CNNs continuously learn and improve their ability to recognize these patterns as they analyze more data. This self-learning capability allows them to achieve high accuracy in disease diagnosis, even for complex conditions like Alzheimer's or cancer.

The benefits of CNNs extend beyond accuracy. They can also analyze large amounts of data much faster than humans, enabling quicker diagnoses and potentially saving lives. This efficiency is particularly valuable in resource-limited settings where timely diagnosis is crucial. Overall, CNNs hold immense potential for revolutionizing disease diagnosis, especially in areas with limited access to specialized healthcare.

LITERATURE REVIEW

The use of machine learning in illness prediction is examined in [1]. Because it makes early identification and intervention possible, disease prediction is essential to healthcare because it improves patient outcomes and lowers costs. The article explores the different ways that machine learning methods, such as using patient data to train predictive models, can be applied to the prediction of disease. These models are able to find patterns and connections linked to particular diseases by analyzing a variety of datasets that include patient demographics, medical histories, genetic data, lifestyle factors, and findings from diagnostic tests. Machine learning algorithms are able

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 5, May 2024

to forecast future disease by learning from past data. events, enabling medical professionals to proactively intervene and create individualized treatment regimens for people who are at risk. The study also probably addresses the drawbacks and difficulties of using machine learning to forecast disease, such as problems with data quality, interpretability of models, and moral issues with patient privacy and permission. All things considered, the study adds to the expanding corpus of research that aims to use machine learning to transform illness prediction and enhance healthcare delivery.

The findings in [2] provide a thorough investigation into the creation of disease prediction models using machine learning techniques. Disease prediction models are essential healthcare tools that help with early diagnosis and treatment of a range of illnesses. The process used to create and apply these prediction models, which usually entails multiple crucial steps, is probably covered by the writers.

They would start by talking about data gathering, which would include finding pertinent datasets that included details on the demographics, medical histories, lifestyle choices, and other relevant elements of the patients. Subsequently, the preprocessing stage would be initiated, whereby methods like feature selection, normalization, and data cleaning might be utilized to guarantee the accuracy and suitability of the data utilized for training the model. The authors would then probably go over how to choose machine learning algorithms that are suitable for tasks involving the prediction of diseases. This could entail investigating different algorithms, including logistic regression, decision trees, support vector machines, and neural networks, or more sophisticated approaches, like ensemble methods and neural networks.

Additionally, the process of training, validating, and evaluating the model would probably be covered in length in the publication. The dataset will be divided into training and testing sets, model hyper-parameters will be adjusted, and the performance of the trained models will be evaluated using suitable metrics including accuracy, precision, recall, and F1-score. The authors may also talk on the models' interpretability and any difficulties that could have arisen during the process of developing the models, such as imbalanced data, overfitting, or problems with generalization.

Furthermore, the study may shed light on the usefulness and possible uses of the created illness prediction models in actual healthcare environments. This could involve talking about patient risk assessment, tailored treatment plans, and the incorporation of predictive analytics technologies into clinical decision support systems. Overall, by demonstrating the efficiency and usefulness of machine learning-based methods in illness prediction and prevention, the work advances healthcare technology.

The insights in [3] offer a thorough examination of computer-aided methods for the diagnosis of skin cancer. Skin cancer is a serious public health issue, and better patient outcomes and successful treatment depend on early detection. The rationale for creating computer-aided skin cancer detection systems is probably covered in the first section of the study, along with the drawbacks of manual diagnosis and the possible advantages of automated methods.

It is likely that the authors will discuss computer-aided detection methods, which may include steps like feature extraction, classification, and image preparation. To improve the consistency and quality of input photos, preprocessing methods including noise removal, contrast enhancement, and image normalization can be utilized. In order to obtain pertinent details about skin lesions that may be symptomatic of malignancy, feature extraction techniques may involve texture analysis, color segmentation, and form descriptors.

The selection and optimisation of machine learning methods for classification problems is probably another topic covered in the study. In order to determine whether a particular algorithm—such as support vector machines, artificial neural networks, or ensemble methods—is effective at differentiating between benign and malignant lesions, it may be necessary to train classifiers using annotated datasets of photos of skin lesions.

The outcomes of tests carried out to assess the effectiveness of the suggested computer-aided detection system may also be presented by the writers. This may involve measuring the system's capacity to accurately identify cases of skin cancer while reducing false positives and false negatives. Relevant metrics such as sensitivity, specificity, and accuracy may be assessed.

The study might also cover the difficulties and practical issues involved in implementing computer-aided detection systems in actual healthcare settings. This could involve talking about things like how to integrate with the current healthcare system, how to comply with regulations, and how to get accepted by medical professionals.

Overall, by presenting the creation and assessment of a computer-aided skin cancer detection system, the work advances the field of medical image analysis. Such technologies may help dermatologists and other healthcare professionals diagnose skin cancer more accurately and quickly by utilizing computational methods and machine learning algorithms, which would eventually improve patient care.

FINDINGS

CNNs are not the primary focus of [1]; rather, it is more likely to concentrate on general machine learning techniques for disease prediction. Below is a summary of the differences between CNNs and every characteristic listed in the paper:

Engineering features:

CNNs' automated feature learning is one of its main advantages. The typical approach employed in the paper requires manual feature engineering, which is a vital step, to extract useful features from the data (pictures). This is not necessary with their approach.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 5, May 2024

Preparing data:
Preprocessing (cleaning, normalization) of the data is necessary for both CNNs and conventional techniques. CNNs may, however, employ preparation techniques, such as picture scaling or normalization, that are designed for image data.

Cross-Checking:
For both methods, cross-validation is necessary to assess model performance and prevent overfitting. There aren't many differences between CNNs and conventional techniques.

Software Tools and Libraries:
CNNs use specific deep learning frameworks like TensorFlow, PyTorch, or Keras, which offer improved functionality for CNN model development and training, while classical approaches may be applied in a variety of machine learning libraries.
Essentially, CNNs automate feature learning, which is a strength, but in contrast to the basic machine learning approach described in the study, they require certain software tools and possibly more powerful hardware.
For the purpose of classifying skin cancer, the traditional approach most likely focuses on developing a machine learning pipeline with manual feature extraction. CNNs, on the other hand, concentrate on automatically extracting features from picture data. This lessens the possibility of bias in the feature selection process and does away with the requirement for human knowledge.

Data Efficiency:
CNNs can be more data-efficient compared to traditional machine learning models. By automatically learning features, they can potentially achieve good performance even with smaller datasets of skin cancer images. This is because the features learned from the data can be highly relevant to the classification task.

Generalizability:
When used with CNNs, data augmentation approaches can increase their generalisability. CNNs can potentially improve classification accuracy on unseen data by learning to be more resilient to small variations in real-world skin cancer photos by the artificial creation of variations of preexisting images.

Changeability:
CNNs are always changing as new pre-trained models and architectures are created. This saves time and effort for researchers by enabling them to take advantage of these developments and maybe enhance the effectiveness of skin cancer categorization.
In general, CNNs are a strong substitute for conventional machine learning techniques in the categorization of skin cancer. They are a useful tool in this field due to their strengths in automatic feature learning, data efficiency, and generalisability; yet, they may need more processing resources and have interpretability restrictions.
Convolutional Neural Networks (CNNs) present a revolutionary method for picture analysis, especially for diagnosing medical conditions like pneumonia. CNNs are exceptionally good in automatically learning and extracting complex visual properties straight from picture data, in contrast to more conventional techniques like Earth Mover's Distance (EMD). This built-in ability enables CNNs to detect minute details such as patterns, textures, and forms that are essential for differentiating between lung images that are healthy and those that are affected by pneumonia. On the other hand, EMD might not be able to capture these subtle traits due to its dependence on comparing probability distributions of pixel intensities, which could restrict its usefulness in intricate medical imaging tasks. CNNs thus represent a promising improvement, providing the possibility for more robust and accurate diagnostic capabilities in healthcare applications, in addition to automated feature learning.

A major influence on Earth Mover's Distance (EMD) performance can come from the number and quality of training data used to create reference distributions for both healthy and pneumonia-infected lungs. By using data augmentation techniques, Convolutional Neural Networks (CNNs) provide an alternative approach to alleviate data reliance. CNNs can expose themselves to a wider range of lung X-ray appearances by artificially increasing the training dataset. This augmentation improves the model's ability to generalize well to new data while also enriching the learning process. As a result, CNNs show that they can reduce the drawbacks caused by reliance on data, which enhances their performance in medical imaging applications such as pneumonia identification.
Convolutional Neural Networks (CNNs) provide solutions to overcome low generalisability, whereas the Earth Mover's Distance (EMD) approach may face difficulties in adjusting to alterations in picture capture methodologies or patient demographics not observed during training. Pre-trained CNN models can be improved for pneumonia recognition tasks by transfer learning. These models were first trained on extensive and varied picture datasets. This method becomes particularly useful when working with restricted medical picture datasets since the model can improve its performance by using pre-learned features from natural photos. Furthermore, the ever-changing field of CNN research makes it easier to continue improving the classification of pneumonia. By using pre-trained models and CNN architectures, researchers can improve classification algorithms over time by iteratively improving their performance. CNNs are positioned as a viable method for enhancing the generalisability and efficacy of pneumonia diagnosis in medical imaging due to their versatility and possibility for ongoing improvement.
Although the EMD method is interpretable, CNNs give a more flexible and comprehensive framework for the detection of pneumonia in chest X-rays. They can attain greater accuracy and possibly more effective generalization to real-world settings because they can automatically acquire pertinent features, make use of data augmentation, and benefit from transfer learning.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 5, May 2024

CONCLUSION

With an emphasis on illnesses such liver problems, hepatitis, heart disease, diabetes, and chronic kidney disease, the study investigates the value of early and precise disease identification for enhancing patient outcomes. The proposed illness identification system attempts to improve illness identification accuracy by automatically learning complex patterns from raw medical images without the need for manual feature extraction. It does this by leveraging Convolutional Neural Networks (CNNs), a potent deep learning technology.

The survey suggests a DDS with a Graphical User Interface (GUI) to enable medical practitioners in effective patient screening and diagnosis using datasets from Kaggle. It looks at the four primary causes of disease: organ failure, infection, deficiency, and genetics. It also highlights how CNNs can identify patterns in different kinds of medical data, such as CT and X-ray scans. According to the poll, CNNs are the best at managing massive amounts of data, getting remarkably accurate results, and speeding up diagnosis—all of which are very useful in situations where resources are scarce. It examines earlier techniques, points out shortcomings in the state-of-the-art methodologies, and highlights the promise of deep learning and sophisticated neural networks to deal with these issues. Much research that emphasizes the combination of machine learning and image processing methods for disease diagnosis—including the identification and prognosis of dermatological disorders—are included in the review of the literature. In order to demonstrate how deep learning techniques, such CNNs, can automate feature extraction and boost diagnostic accuracy, applications such as brain MRI segmentation and skin cancer classification are investigated. CNNs and traditional machine learning methods are compared in the research gap study, with a focus on CNNs' automated feature learning, data efficiency, and generalisability. CNNs show promise in medical picture processing and could lead to improvements in illness diagnosis, especially in complicated cases like pneumonia. Conclusively, the survey and literature analysis underscore the revolutionary possibilities of CNNs and deep learning in the diagnosis of diseases, providing valuable perspectives for future research paths targeted at enhancing patient outcomes and healthcare provision.

REFERENCES

[1] P. S. Kohli and S. Arora, “Application of machine learning in disease prediction,” in 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1–4.

[2] D. Dahiwade, G. Patle, and E. Meshram, “Designing disease prediction model using machine learning approach,” Proceedings of the 3rd International Conference on Computing Methodologies and Communication,ICCMC 2019, no. ICCMC, pp. 1211–1215, 2019.

[3] Khatri A, Jain R, Vashista H, et al. Pneumonia identification in chest X-ray images using EMD. Singapore: Springer; 2020.

[4] Ahsan M.M., Gupta K.D., Islam M.M., Sen S., Rahman M., Shakhawat Hossain M. COVID-19 symptoms detection based on nasnetmobile with explainable ai using various imaging modalities. Mach. Learn. Knowl. Extr. 2020;2:490–504. doi: 10.3390/ make2040027.

[5] Monika, M. & Vignesh, N. & Kumari, Usha & Kumar, M.N.V.S.S. & Lydia, Laxmi. (2020). Skin cancer detection and classification using machine learning. Materials Today: Proceedings. 33. 10.1016/j.matpr.2020.07.366

[6] A Sharma et al 2021 IOP Conference. Ser.: Mater. Sci. Eng. 1022 012066.