Journal of Advanced Health Informatics Research (JAHIR)
Vol. 2, No. 1, (2024), pp. 50-63
ISSN: 2985-6124, DOI: 10.59247/jahir.v2i1.195

Predicting Internal Diseases in Humans Using Machine Learning: A Systematic Literature Review

Rosyid R. Al-Hakim [1], Yurii Prokopchuk [2]
[1] *Primatology Study Program, Graduate School, IPB University, Jl. Lodaya II No.5, Bogor 16151, Indonesia*
[2] *Department of Computer Science, Information Technologies and Applied Mathematics, Chernyshevsky Street 24a, Dnipro 49600, Ukraine*

ARTICLE INFO

Article history:

Received March 12, 2024
Revised January 01, 2025
Published January 26, 2025

Keywords:

Artificial Intelligence;
Classification;
Internal Medicine;
Literature Review;
Machine Learning.

ABSTRACT

Human health is the main focus of clinical medicine, especially in understanding internal diseases involving the body's organs. Identifying and predicting disease at an early stage is essential to prevent the development of more severe disease. These challenges encourage using the latest technologies, especially machine learning techniques. This technology is used to ensure accurate disease predictions. The results of the research identified various types of internal diseases, including heart, kidney, lung and liver cancer, and highlighted the associated symptoms and risk factors. Several algorithms are used to classify internal diseases, including the classification of heart disease. The logistic regression algorithm is the most efficient, with accuracy results of 88.52%. Meanwhile, CHIRP kidney disease classification provides the most efficient results with an accuracy of 99.75%. MobileLungNetV2 has an accuracy of 96.97% for lung disease classification, and classification for liver disease produces the highest accuracy in logistic regression at 72.50%. Machine learning in disease prediction significantly contributes, especially in increasing accuracy and efficiency in diagnosis and risk prediction. Despite significant progress, challenges such as dataset size, data quality, and model validation need to be addressed to maximise the potential of this technology in clinical practice.

This work is licensed under a Creative Commons Attribution-Share Alike 4.0

Corresponding Author:

Rosyid R. Al-Hakim, Primatology Study Program, Graduate School, IPB University, Jl. Lodaya II No.5, Bogor 16151, Indonesia
Email: alhakimrosyid@apps.ipb.ac.id

1. INTRODUCTION

Clinical medicine studies various aspects of internal disease, including diseases of internal organs, how they spread, their causes, development processes, classification, clinical symptoms, methods of diagnosis and comparison with other diseases, preventive measures, and appropriate treatment [1]. Internal medicine refers to various types of diseases that affect the organs in the human body, such as diabetes mellitus, hypertension, heart failure, and kidney disease [2].

In the results and discussion, the research focused on heart, liver, lung and kidney diseases. Other diseases, such as diabetes mellitus and hypertension, are not discussed within the framework of this research. This focus selection was made to narrow the scope of analysis and pay special attention to certain aspects of clinical medicine. Nevertheless, this research still contributes to understanding and managing significant internal diseases.

Identifying and predicting these diseases in early conditions cannot be underestimated, as it can prevent the development more severe diseases. The challenge lies in the difficulty for doctors to identify diseases with high accuracy over time manually. Besides, to identify and predict chronic diseases in patients, the latest machine learning techniques can ensure reliable categorisation in identifying individuals suffering from chronic diseases [3].

Machine Learning combines various interdisciplinary scientific disciplines, such as logic, statistics, and probability theory. According to Arthur Samuel, machine learning is a field of study that allows computers to learn without requiring explicit programming. Machine learning (ML) teaches machines to handle data more efficiently [4]. One of the advantages of machine learning is the ability to extract potential information from vast and complex data and apply it for decision-making. Machine learning can solve problems such as collinearity and multivariate interactions because of its scalable and flexible nature [5].

Machine learning describes an environment that aligns science, informatics, incentives, and culture for continuous improvement and innovation. Combining diverse data sources with complex machine learning algorithms would provide continuous data-based insights to optimise biomedical research public health and improve healthcare quality[6]. In health, machine learning is used to determine or extract medical knowledge, opening up new ideas for practitioners and specialists. In clinical practice, machine learning predictive models can highlight improved rules for patient care decision-making[7].

Disease prediction using machine learning is a system used to predict disease from symptoms provided by patients or users. The system processes symptoms the user provides as input and provides output in the form of disease probability [8]. A significant challenge in developing early diagnostic tools and effective treatment is the complexity of different disease mechanisms and underlying symptoms in patient populations. One reliable solution is the application of machine learning, which is part of the field of artificial intelligence (AI). Using machine learning, researchers, doctors, and patients can overcome several problems related to this complexity [9].

2. METHODS

This research uses a literature review study. The literature study search strategy uses online databases Science Direct, Google Scholar, and the Institute of Electrical and Electronics Engineers (IEEE) with machine learning, prediction, classification, and internal medicine. The articles are written in Indonesian and published in the last 5 years.

Using specific characteristics, researchers filtered the 130 papers found through a literature search regarding machine learning, prediction, classification, and internal medicine. The selection process is carried out carefully, considering the relevance and quality of each article. These characteristics include aspects such as the machine learning method approach used, the accuracy of research results, and the unique contribution each selected article made. As a result, several selected references that best suited the objectives of this research were selected for further analysis.

3. PREVIOUS WORKS

A brief overview of studies using machine learning to predict internal diseases. Studies already use machine learning, as seen in Table 1.

Table 1. Related Studies Used for Review

No	Reference	Author(s)	Year	Results
1.	Heart Disease Prediction Using Machine Learning Techniques	Venu Gopal, T. Sneha, Avudurthi [10]	2023	Using machine learning algorithms, especially Gaussian Naive Bayes, can provide high accuracy in predicting heart disease. Its advantages are the ability to provide accurate predictions and the potential to detect heart disease early. However, the downside is limitations in existing systems and the need for better analysis to improve prediction accuracy. In addition, it is crucial to detect heart disease early and suggest further development

				using machine learning approaches for better analysis.
2.	Heart disease prediction using machine learning algorithms	Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain, Preeti Nagrath [11]	2021	Machine learning algorithms, such as logistic regression and KNN, can predict and classify patients with heart disease. The model accuracy achieved was 87.5%, with KNN being the most accurate algorithm with an accuracy of 88.52%. This model can help predict patients with heart disease and improve medical care. The advantage of this research is the ability of machine learning algorithms to predict heart disease with high accuracy, which can help in early diagnosis and patient treatment. In addition, the use of machine learning techniques can also help in reducing the costs of medical care by enabling the identification of patients with heart disease more efficiently. However, the shortcomings of this research are limitations in the data used and dependence on the quality and representation of the available data.
3.	Chronic kidney disease prediction based on machine learning algorithms	Md. Ariful Islam, Md. Ziaul Hasan Majumder, Md. Alomgeer Hussein [12]	2023	Use of machine learning algorithms to diagnose chronic kidney disease early. The study used a dataset from the UCI Machine Learning Repository, applied predictive modelling, and used Principal Component Analysis (PCA) to reduce the number of input features. The results showed positive and negative correlations between several features and chronic kidney disease. Advantages include the use of a variety of different machine learning algorithms and data processing to handle missing values and categorical variables. However, a drawback is the small size of the dataset, which may affect the model's accuracy.
4.	Prediction of Chronic Kidney Disease - A Machine Learning Perspective	Pankaj Chittora, Sandeep Chaurasia, Prasun Chakrabarti, Gaurav Kumawat, Tulika Chakrabarti, Zbigniew Leonowicz, Michał Jasiński, Łukasz Jasiński, Radomir Gono, Elżbieta Jasińska, Vadim Bolshev [13]	2021	This research aims to predict chronic kidney disease using machine learning models and artificial neural networks. The results show that the highest accuracy achieved is 99.6% using various classification algorithms and class balance techniques. The best machine learning model is a Linear Support Vector Machine (LSVM) with an L2 penalty and Lambda 0.5. The methodological process includes data pre-processing, feature selection, application of classification algorithms, use of SMOTE techniques, and statistical analysis to compare results from machine learning models and artificial neural networks. The advantage is that the highest accuracy is 99.6%, showing the model's ability to

				predict chronic kidney disease very well. Using SMOTE for class balance improves the model's ability to handle class imbalance in the dataset. The drawback is that a more extensive dataset size might provide a more comprehensive understanding of chronic kidney disease prediction.
5.	Machine Learning and Feature Selection Methods for Disease Classification with Application to Lung Cancer Screening Image Data	Darcie AP Delzell, Sara Magnuson, Tabitha Peter, Michelle Smith, Brian J Smith [14]	2019	Radiomics is an approach to extracting quantitative data from medical images, such as CT scans, to identify prognostic biomarkers and tumour classification. Radiomics has effectively identified prognostic biomarkers for lung, head and neck cancer. The radiomics approach involves extracting features from medical images, such as CT scans, using techniques such as Laws' Texture Energy Measures (TEM) to extract texture features. These features are then used for classification and prediction in machine learning models, such as support vector machines. The advantage is that Radiomics with machine learning promises to be an effective diagnostic tool for tumour classification, with the potential to provide good classification and simultaneously reduce false favourable rates; this approach has also been proven effective in identifying prognostic biomarkers for various types of cancer. The drawback is that some machine learning models, such as random forests and bagged trees, have proven to be less effective in classification using radiomic biomarkers. Additionally, there are challenges in precise radiomic feature extraction and model validation that require extensive and diverse data.
6.	Digital Technology and the Future of Interstitial Lung Diseases 2	Hayley Barnes, Stephen M Humphries, Peter M George, Deborah Assayag, Ian Glaspole, John A Mackintosh, Tamera J Corte, Marilyn Glassberg, Kerri A Johannson, Lucio Calandriello, Federico Felder, Athol Wells, Simon Walsh [15]	2023	Use of machine learning in the diagnosis and prognosis of interstitial lung disease. The methods used include quantitative CT analysis and the application of machine learning in medical image biomarker research, including radiomics, deep learning, and quantitative CT analysis in diagnosing and monitoring lung disease. The results include improvements in the efficiency, precision, and reproducibility of segmentation methods in medical images, as well as the potential of this advanced technology in improving the diagnosis and management of lung diseases. Its advantages include the ability to predict disease progression, assess lung disease severity, and predict patient outcomes.

				The drawback is the need for proper validation, interpretation, and implementation in clinical practice.
7.	XRAY AI: Lung Disease Prediction Using Machine Learning	Justin Monsi, Justine Saji, Keerthy Vinod, Liya Joy, Jis Joe Mathew [16]	2019	Uses an NIH X-ray dataset covering 14 chest diseases. The process involves data pre-processing, model training, and model testing. The results show that this automatic detection system can be integrated with hospital management systems and used as a doctor verification tool. The system's accuracy can be improved by adding attributes such as smoking history and family history. The lack of a large enough dataset is a problem in model training, mainly when identifying lung nodules and lung cancer. The researchers suggested that the model performance tended to have low sensitivity overall due to the limited dataset. Besides, this means the model may be less responsive in detecting conditions it could identify more accurately.
8.	Artificial intelligence in liver cancers: Decoding the impact of machine learning models in clinical diagnosis of primary liver cancers and liver cancer metastases	Anita Bakrania, Narottam Joshi, Xun Zhao, Gang Zheng, Mamatha Bhat [17]	2023	Using various methods such as deep learning, convolutional neural networks (CNN), logistic regression, random forest, support vector machine (SVM), and other algorithms. The results show that the use of artificial intelligence (AI) in diagnosing and managing liver cancer, including hepatocellular carcinoma (HCC), Chol angiosarcoma (CCA), and metastatic colorectal liver cancer, shows promising results. AI has been used to differentiate between HCC and CCA, classify HCC tumours, and predict early recurrence in CCA patients. The use of AI in diagnosing and managing liver cancer offers the potential to improve pre-screening analysis, diagnosis and prognosis of the disease. Additionally, AI can also improve personalised therapy and clinical efficiency. The disadvantages are that AI can also experience underfitting or overfitting data, has a bias towards more general classes, requires external validation to ensure model generalisability, has extensive data for model training, and is dependent on intensive hardware and long training times. long
9.	Theragnostic role of machine learning in clinical management of kidney stone disease	Supatcha Sassanarakkit, Sudarat Hadpech, Visith Thongboonkerd [18]	2022	Uses various machine learning methods, including achieved through the use of machine learning in the paper, including prediction of stone-free status after PNL for staghorn stones with an accuracy of 81%. Apart from that, there is also a prediction of kidney stone composition with an accuracy of 98.17%.

				Although machine learning methods have provided significant results, there is still room for improvement of machine learning algorithms to increase the sensitivity and specificity of automatic classification methods, especially for ureteroscopy kidney stone images. In addition, blood and urine chemistry laboratory tests also need to be combined with clinical information and medical images to improve the accuracy of machine learning in KSD theragnostic.
10.	Prediction of pancreatic cancer risk in patients with new-onset diabetes using a machine learning approach based on routine biochemical parameters	Simon Lebech Cichosz, Morten Hasselstrøm Jensen, Ole Hejlesen, Stine Dam Henriksen, Asbjørn Mohr Drewes [19]	2023	Developing a machine-learning model to predict the risk of pancreatic cancer in individuals with new-onset diabetes. The advantage is that this model has good discrimination performance with an AUC-ROC of 0.78. Age and rate of change in HbA1c were the most significant differentiators. This model was able to predict the risk of PCRD with a 20-fold increase in relative risk compared with the general population of new-onset diabetes. However, there is a lack of external validation cohorts and a limitation of the PCRD cohort to only individuals with HbA1c information. Additionally, this model has limitations in a population-based setting and PCRD definitions based on ICD-10 codes without histological and clinical case verification.

Collect, evaluate, and synthesise previous research results using literature analysis or literature review techniques. In this process, an interdisciplinary and multilevel approach is used. The result, namely selecting appropriate machine learning methods and models for each disease, shows the diversity of approaches and advances in using this technology in the medical field. Machine learning in internal medicine prediction has shown promising results in diagnosing and managing heart, kidney, lung and liver cancer diseases. Machine learning algorithms can provide high-accuracy disease predictions, helping in early diagnosis, treatment and reduction of medical care costs. However, there are challenges in dataset size, data quality, model validation, and bias in medical data that need to be addressed. Further research is needed to realise the potential of machine learning in healthcare fully.

4. RESULTS AND DISCUSSION

4.1. Identify Internal Diseases

There are several types of internal diseases that millions of people suffer from, and these diseases can disrupt the body's internal organs. The following are some of the most common internal diseases.

Heart Disease

Heart disease is one of the leading causes of death worldwide. Common symptoms of heart disease include shortness of breath, physical fatigue, and swollen legs [20]. People in their forties and fifties are more susceptible to heart disease, with several factors such as personal lifestyle, medical history, and genetic predisposition playing an essential role in the development of this disease. The risk of heart disease may increase due to factors such as smoking, excessive alcohol consumption, an inactive lifestyle, depression, obesity, chronic stress, genetic disorders, and pre-existing heart problems [21]. Electronic Health Records (EHR) are tools frequently used to discover valuable data patterns that improve the predictions of machine learning algorithms. In particular, machine learning is very helpful in solving problems such as prediction in areas such as healthcare [22].

Kidney Disease

The kidneys are vital in regulating the amount of water and minerals in the blood and maintaining the body's internal stability. Impaired kidney function can harm other parts of the body [23]. Chronic kidney disease is a clinical syndrome caused by significant kidney function and structure changes. These changes are irreversible, and the condition develops slowly and progressively. A higher risk of complications and death, especially those related to the cardiovascular system, is another critical factor [24].

Routine examinations with blood and urine tests are often used to detect chronic kidney disease. Some symptoms may include blood-coloured or foamy urine, frequent urination at night, pelvic pain, or decreased urine output. Patients may also experience a lack of appetite, nausea, vomiting, metallic taste, weight loss for no apparent reason, itching, mental changes, shortness of breath, or swelling of the legs if the kidney disease is severe [25].

Lung Cancer

The lungs consist of several essential parts, including the bronchi, small bronchi, alveolar tubes, alveoli, and pulmonary blood vessels. The pulmonary vasculature includes the pulmonary arteries, veins, and microvessels. The lungs also have an alveolar surface area, essential in gas exchange [26]. According to the World Health Organization (WHO), more than 4 million premature deaths each year are related to diseases caused by indoor air pollution, such as asthma and pneumonia. This situation shows that protecting lung health is an important priority, especially in areas that are vulnerable to these risk factors [27].

Due to its high malignancy and rapid progression, lung cancer is the second most frequently diagnosed cancer worldwide and is the leading cause of cancer death. More than 70% of lung cancer patients are diagnosed at advanced or locally advanced stages 3, 4 or 5 and have no chance of undergoing surgery due to the absence of specific symptoms in the early stages. These patients typically receive radiation, chemotherapy, and molecular-targeted therapy [28].

Liver Disease

Liver disease causes approximately 2 million deaths each year worldwide: 1 million due to complications of cirrhosis and 1 million due to viral hepatitis and hepatocellular carcinoma. Approximately 2 billion people consume alcohol worldwide and are at risk of alcohol-related liver disease. About 2 billion adults are overweight or obese, and more than 400 million have diabetes, both risk factors for non-alcoholic fatty liver disease (NAFLD) and hepatocellular carcinoma [29].

The most common chronic liver disease is non-alcoholic fatty liver disease (NAFLD) [30]. Symptoms of NAFLD range from increased intrahepatic lipid content (steatosis, non-alcoholic fatty liver, or NAFL) to non-alcoholic steatohepatitis (NASH), which is accompanied by varying degrees of fibrosis, necrotic inflammation, and ultimately cirrhosis. NAFLD is associated with a higher risk of hepatocellular carcinoma, as well as cardiovascular disease and complications associated with diabetes mellitus, such as nephropathy and neuropathy [31].

4.2. Contribution of Machine Learning to Predicting Internal Diseases in Human

Machine Learning in Heart Disease Classification

The medical industry needs high-performance models to accurately predict a patient's likelihood of developing heart disease [32]. Machine learning can be beneficial in determining locomotor disorders, heart disease, and other conditions [33]. Machine learning is widely used in cardiac imaging as the number of Coronary Computed Tomography Angiography (CCTA) examinations increases. ML algorithms can provide better prediction results than traditional risk scores using imaging reports and clinical parameters [34]. Figure 1 is a process for predicting heart disease.

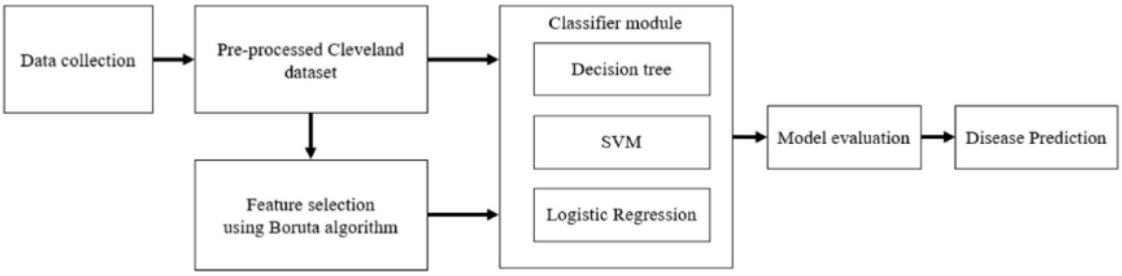

Fig. 1. Steps to predict heart disease.

Figure 1 [35] explains the system that uses the Boruta feature selection algorithm to obtain relevant features from the Cleveland Clinic Heart Disease dataset obtained from Kaggle. The first stage is feature selection, where the Boruta algorithm determines the features that influence prediction accuracy the most. The second stage is creating a prediction model using machine learning algorithms such as decision trees, support vector machines (SVM), and logistic regression. These models would be trained using data from the dataset for which the features have been selected. The final stage is the evaluation of model performance. In this stage, evaluation metrics such as the confusion matrix measure how much these models can accurately predict heart disease.

The results obtained are that the Decision Tree (DT) has an accuracy of 75.41%, the Support Vector Machine (SVM) has an accuracy of 81.97%, and the Logistic Regression (LR) has an accuracy of 88.52%. Based on these results, it can be concluded that logistic regression provides the most efficient results [35].

Machine Learning in Kidney Disease Classification

Conditions that can cause kidney failure include diabetes, hypertension and heart problems. Age and gender can also play a role in determining the risk of kidney failure. If one or both kidneys are not functioning correctly, symptoms may appear, such as back pain, stomach pain, diarrhoea, fever, nosebleeds, rash, and vomiting [36].

Kidney disease is a long-term illness in which the kidneys gradually lose their function, sometimes leading to kidney failure. The disease usually does not show symptoms in its early stages, so many people cannot detect it. Therefore, detection of kidney disease is only possible in advanced stages when many symptoms are present [37].

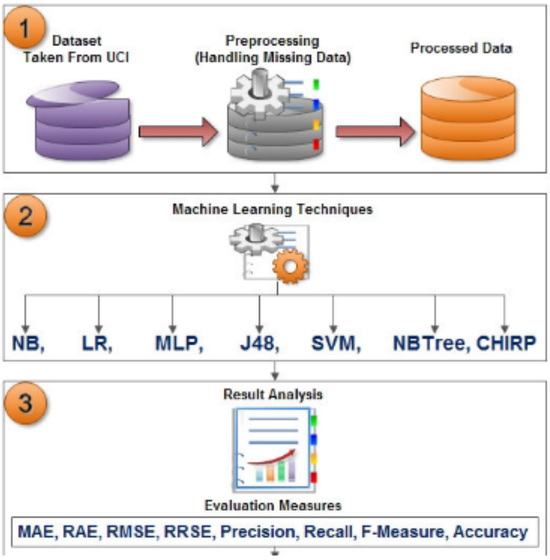

Fig. 2. Steps to predict kidney disease.

Figure 2. [38] is a step in predicting kidney disease by pre-processing the data because some values are missing in the dataset taken from the UCI ML repository. Missing values are addressed by taking the average of the existing values and replacing the missing fields with the average value. After this step, ML techniques are applied to the dataset, and the results are analysed using accuracy and error rates.

There are seven algorithms used, namely, CHIRP has an accuracy of 99.75%, NBTree (Naive Bayes Tree) with an accuracy of 98.75%, SVM (Support Vector Machine) with an accuracy of 98.25%, J48 (C4.5 Decision Tree) with an accuracy of 97.75%, MLP (Multilayer Perceptron) with 97.25% accuracy, LR (Logistic Regression) with 96.50% accuracy, and NB (Naive Bayes) with 95.75% of accuracy. Based on these results, it can be concluded that CHIRP provides the most efficient results [38].

Machine Learning in Lung Disease Classification

About 2.20 million new patients are diagnosed with lung cancer every year, and 75% of them die within five years of diagnosis. It is one of the most commonly diagnosed cancers and a cause of cancer deaths worldwide [39]; with advances in image processing and analysis, machine learning algorithms can identify chronic lung diseases at risk, predict the degree of pulmonary fibrosis, find associations between radiological abnormalities and decreased lung function, and be used as endpoints in treatment trials, showing how this technology can be used in the treatment of chronic lung disease sufferers [40].

Machine learning can also help doctors make decisions. Computed Tomography (CT) and Fluorodeoxyglucose Positron Emission Tomography (FDG-PET) are two types of imaging commonly used to diagnose lung cancer. Combining the two provides essential information for predicting prognosis and guiding treatment choices. Patients at low risk based on FDG-PET/CT may choose invasive procedures for cure, while patients at high risk may consider chemotherapy, radiation, targeted therapy, or immunotherapy [41]. Figure 3 [42] explains the steps for classifying lung disease.

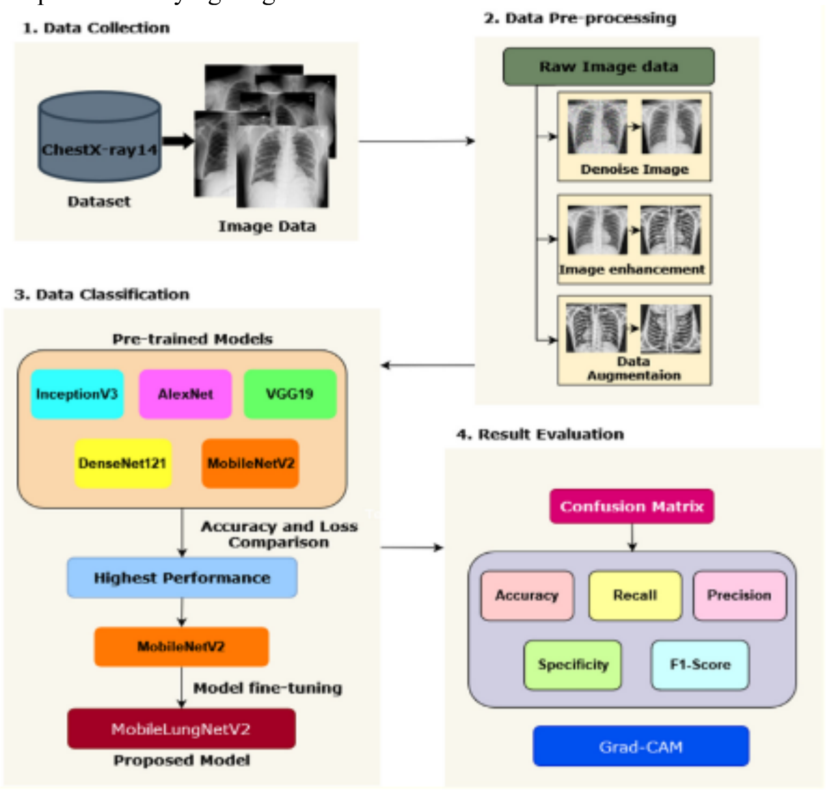

Fig. 3. Steps to classifying lung disease.

In classifying lung disease, several steps are taken, namely the first is collecting data in the form of datasets and image data; the second is image pre-processing; the third is classifying the data; and the fourth is displaying the evaluation results. We have obtained results from these steps showing that the

MobileLungNetV2 model has a lung lesion classification accuracy of 96.97% after fine-tuning. Furthermore, this model has high specificity, recall, and f1-score for most lung lesions [42].

Machine Learning in Liver Disease Classification

The liver is the largest organ in the body, which helps digest food and remove toxic elements from the body. There are many types of liver disease, such as hepatitis, cirrhosis, liver tumours and liver cancer. Liver disease can be caused by obesity, undiagnosed hepatitis infection, and alcohol abuse [43]. Machine learning in liver disease prediction can identify patterns and trends in patient data, which can help predict patient outcomes, such as length of hospital stay after liver transplantation [44]. Figure 4 [45] explains the flow of liver disease classification.

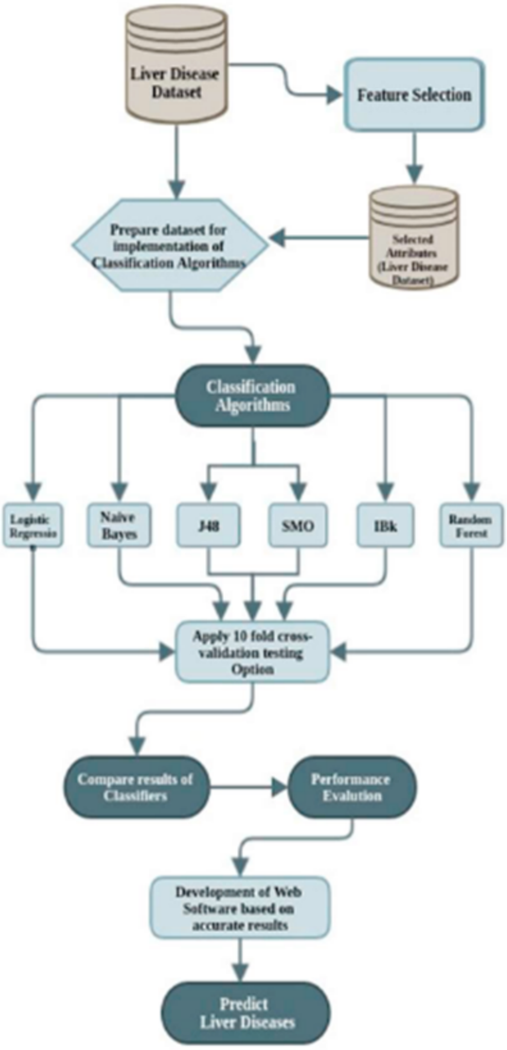

Fig. 4. Flow of liver disease classification.

The steps in classifying liver disease include, first, database selection by preparing a dataset of liver patients in the ARFF (Attribute Relation File Format) file format; then, the data is selected by deleting unnecessary fields, missing records, and duplicate records if any, then the data that has been selected would be processed and transformed and then the data would be classified using several algorithms, namely Logistic Regression, Naive-Bayes, J48, SMO, IBk, and Random Forest. Besides, to compare the results of the various classification algorithms, the performance of each classifier is evaluated based on the correctly classified instances and execution time.

The results obtained from each algorithm are that Logistic Regression has an accuracy of 72.50%, Naive-Bayes has an accuracy of 55.74%, SMO has an accuracy of 71.35%, IBk has an accuracy of 64.15%, has an accuracy of 68.78% and Random Forest has an accuracy of 71.53%. Based on these results, it can be concluded that Logistic Regression provides the most efficient results [45].

4.3. The Advantages of Machine Learning in Health Data Analysis

Machine learning has advantages in health data analysis because it can predict disease with a data-based approach. Another advantage is Machine Learning's ability to process big data to find non-linear relationships in electronic medical records [46]. ML algorithms can analyse electronic health records, imaging, wearable devices, and sensor data, provide personalised risk assessments and promote guideline-compliant medical management [47].

4.4. Disease Risk Prediction with Machine Learning

Existing risk prediction algorithms are typically developed using multivariate regression models that combine information on several well-established risk factors and generally assume that all of these factors are related to CVD outcomes in a linear manner, with limited or no interaction between different factors. Data-driven techniques based on machine learning (ML) can improve risk prediction performance by agnostically discovering new risk predictors and learning the complex interactions between them [48].

4.5. Advantages of Using Machine Learning in the World of Health

The benefits of machine learning include flexibility and scalability compared to conventional biostatistical methods, making it applicable to various tasks such as risk clustering, diagnosis and classification, and survival prediction. Another advantage of machine learning algorithms is their ability to analyse various types of data (such as demographic data, laboratory test results, image data, and doctor's text-free notes) and combine them in predictions of disease risk, diagnosis, prognosis, and appropriate treatment [49].

4.6. Challenges of Using Machine Learning in the Health Sector

Implementing machine learning in healthcare involves the need for high-quality data that can accurately reflect the diversity of populations, to which machine learning models would later generalise. The better the dataset collected, the more it would reflect the diversity of the population, so the more reliable and the more relevant the results produced by the ML model are for practical applications in the real world [50]. Apart from that, the existence of bias in medical data is also a big challenge because it can reduce the reliability of the model, especially if the bias is not corrected during model development, then the resulting machine learning (ML) model can produce inaccurate or unreliable predictions [51]. Besides, to overcome this challenge, efforts must be made to improve data quality by ensuring that medical information is complete and accurate. Additionally, it is essential to address the imbalance and heterogeneity of data populations and detect and address bias through systematic analysis.

4.7. Further Research Potentials

Further research could focus on developing more complex and sophisticated machine-learning models for internal disease prediction. Besides this, it includes using deep learning techniques, ensemble learning, and combining various types of medical data to improve prediction accuracy. Further studies could focus on validating machine learning models in real clinical contexts. In addition, this involves the interpretation of prediction results and the integration of the model into clinical practice to ensure the usefulness and reliability of the predictions. Research can also focus on developing methods for managing and cleaning extensive, complex medical data. Besides, this includes identifying and addressing bias in data and using techniques to improve the quality of medical data used in machine learning models.

5. CONCLUSION

Based on the literature studies, machine learning in predicting internal diseases such as heart, kidney, lung and liver cancer shows promising results in diagnosing and managing these diseases. Machine learning algorithms can provide high-accuracy disease predictions, helping in early diagnosis, treatment and reduction of medical care costs. However, challenges regarding dataset size, data quality, model validation, and the presence of bias in medical data need to be addressed. However, using machine learning in the health sector has excellent potential to provide personalised predictions and support appropriate medical management.

Further research is needed to overcome these challenges and fully realise the potential of machine learning in disease prediction and diagnosis.

REFERENCES

[1] I. Serdaliyeva, A. Omirbayeva, N. Nurmanova, K. Kenzheyeva, and G. Izzatullayeva, "Propedeutics and Problems of Internal Diseases,"*Acta Informatica Medica*, vol. 31, no. 4, p. 287, 2023, doi: 10.5455/aim.2023.31.287-291.

[2] V. Suzan and H. Yavuzer, "A Fuzzy Dematel Method To Evaluate The Most Common Diseases In Internal Medicine,"*International Journal of Fuzzy Systems*, vol. 22, no. 7, pp. 2385–2395, Oct. 2020, doi: 10.1007/s40815-020-00921-x.

[3] R. Alanazi, "Identification and Prediction of Chronic Diseases Using Machine Learning Approach,"*J Healthc Eng*, vol. 2022, 2022, doi: 10.1155/2022/2826127.

[4] B. Mahesh, "Machine Learning Algorithms-A Review," *International Journal of Science and Research*, 2020, doi: 10.21275/ART20203995.

[5] Y. Wu, L. Li, B. Xin, Q. Hu, X. Dong, and Z. Li, "Application of machine learning in personalised medicine," *Intelligent Pharmacy*, vol. 1, no. 3, pp. 152–156, Oct. 2023, doi: 10.1016/j.ipha.2023.06.004.

[6] JAM Sidey-Gibbons and CJ Sidey-Gibbons, "Machine learning in medicine: a practical introduction," *BMC Med Res Methodol*, vol. 19, no. 1, March. 2019, doi: 10.1186/s12874-019-0681-4.

[7] G. Battineni, GG Sagaro, N. Chinatalapudi, and F. Amenta, "Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis," *J Pers Med*, vol. 10, no. 2, p. 21, March. 2020, doi: 10.3390/jpm10020021.

[8] CK Gomathy and A. Rohith Naidu Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, "The Prediction Of Disease Using Machine Learning," *International Journal of Scientific Research in Engineering and Management*, 2021, [Online]. Available: www.ijsrem.com

[9] [9] MM Ahsan, SA Luna, and Z. Siddique, "Machine-Learning-Based Disease Diagnosis: A Comprehensive Review," *Healthcare (Switzerland)*, vol. 10, no. 3. MDPI, Mar. 01, 2022. doi: 10.3390/healthcare10030541.

[10] [10] T. Venu Gopal and A. Sneha, "Heart Disease Prediction Using Machine Learning Techniques Email: drtv.gopal@jntuh.ac.in ," vol. 13, p. 7, 2023.

[11] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," in*IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, Jan. 2021. doi: 10.1088/1757-899X/1022/1/012072.

[12] MA Islam, MZH Majumder, and MA Hussein, "Chronic kidney disease prediction based on machine learning algorithms,"*J Pathol Inform*, vol. 14, Jan. 2023, doi: 10.1016/j.jpi.2023.100189.

[13] P. Chittora*et al.*, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," IEEE Access, vol. 9. Institute of Electrical and Electronics Engineers Inc., pp. 17312–17334, 2021. doi: 10.1109/ACCESS.2021.3053763.

[14] DAP Delzell, S. Magnuson, T. Peter, M. Smith, and BJ Smith, "Machine Learning and Feature Selection Methods for Disease Classification With Application to Lung Cancer Screening Image Data," *Front Oncol*, vol. 9, Dec. 2019, doi: 10.3389/fonc.2019.01393.

[15] H. Barnes*et al.*, "Machine learning in radiology: the new frontier in interstitial lung diseases," The Lancet Digital Health, vol. 5, no. 1. Elsevier Ltd, pp. e41–e50, Jan. 01, 2023. doi: 10.1016/S2589-7500(22)00230-8.

[16] J. Monsi, J. Saji, K. Vinod, L. Joy, and JJ Mathew, "XRAY AI: Lung Disease Prediction Using Machine Learning," *International Journal of Information Systems and Computer Sciences*, vol. 8, no. 2, pp. 51–54, Apr. 2019, doi: 10.30534/ijiscs/2019/12822019.

[17] A. Bakrania, N. Joshi, X. Zhao, G. Zheng, and M. Bhat, "Artificial intelligence in liver cancers: Decoding the impact of machine learning models in clinical diagnosis of primary liver cancers and liver cancer metastases," *Pharmacol Res*, vol. 189, March. 2023, doi: 10.1016/j.phrs.2023.106706.

[18] S. Sassanarakkit, S. Hadpech, and V. Thongboonkerd, "Theranostic role of machine learning in clinical management of kidney stone disease,"*Comput Struct Biotechnol J*, vol. 21, pp. 260–266, 2022, doi: 10.1016/j.csbj.2022.12.004.

[19] SL Cichosz, MH Jensen, O. Hejlesen, SD Henriksen, AM Drewes, and SS Olesen, "Prediction of pancreatic cancer risk in patients with new-onset diabetes using a machine learning approach based on routine biochemical parameters; Prediction of Pancreatic Cancer Risk in New Onset Diabetes," *Comput Methods Programs Biomed*, vol. 244, Feb. 2024, doi: 10.1016/j.cmpb.2023.107965.

[20] JP Li, AU Haq, SU Din, J. Khan, A. Khan, and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare,"*IEEE Access*, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[21] G. Manikandan, B. Pragadeesh, V. Manojkumar, AL Karthikeyan, R. Manikandan, and AH Gandomi, "Classification models combined with Boruta feature selection for heart disease prediction," *Inform Med Unlocked*, vol. 44, p. 101442, 2023, doi: 10.1016/j.imu.2023.101442.

[22] S. Mohapatra, S. Maneesha, PK Patra, and S. Mohanty, "Heart Diseases Prediction based on Stacking Classifiers Model,"*Procedia Comput Sci*, vol. 218, pp. 1621–1630, 2023, doi: 10.1016/j.procs.2023.01.140.

[23] J. Qezelbash-Chamak, S. Badamchizadeh, K. Eshghi, and Y. Asadi, "A survey of machine learning in kidney disease diagnosis," *Machine Learning with Applications*, vol. 10, p. 100418, Dec. 2022, doi: 10.1016/j.mlwa.2022.100418.
[24] [24] AL Ammirati, "Chronic Kidney Disease,"*Rev Assoc Med Bras*, vol. 66, no. suppl 1, pp. s03–s09, 2020, doi: 10.1590/1806-9282.66.s1.3.
[25] TK Chen, DH Knicely, and ME Grams, "Chronic Kidney Disease Diagnosis and Management," *JAMA*, vol. 322, no. 13, p. 1294, Oct. 2019, doi: 10.1001/jama.2019.14745.
[26] W. Zhong, X. Zhang, Y. Zeng, D. Lin, and J. Wu, "Recent applications and strategies in nanotechnology for lung diseases," *Nano Res*, vol. 14, no. 7, pp. 2067–2089, Jul. 2021, doi: 10.1007/s12274-020-3180-3.
[27] S. Bharati, P. Podder, and MRH Mondal, "Hybrid deep learning for detecting lung diseases from X-ray images," *Inform Med Unlocked*, vol. 20, p. 100391, 2020, doi: 10.1016/j.imu.2020.100391.
[28] C. Kong*et al.*, "Machine Learning Classifier for Preoperative Prediction of Early Recurrence After Bronchial Arterial Chemoembolization Treatment in Lung Cancer Patients," Acad Radiol, vol. 30, no. 12, pp. 2880–2893, Dec. 2023, doi: 10.1016/j.acra.2023.04.011.
[29] SK Asrani, H. Devarbhavi, J. Eaton, and PS Kamath, "Burden of liver diseases in the world," *Journal of Hepatology*, vol. 70, no. 1. Elsevier B.V., pp. 151–171, Jan. 01, 2019. doi: 10.1016/j.jhep.2018.09.014.
[30] [30] R. Loomba, SL Friedman, and GI Shulman, "Mechanisms and disease consequences of non-alcoholic fatty liver disease," *Cell*, vol. 184, no. 10, pp. 2537–2564, May 2021, doi: 10.1016/j.cell.2021.04.015.
[31] K. Pafili and M. Roden, "Non-alcoholic fatty liver disease (NAFLD) from pathogenesis to treatment concepts in humans,"*Mol Metab*, vol. 50, p. 101122, Aug. 2021, doi: 10.1016/j.molmet.2020.101122.
[32] PS Asih, Y. Azhar, GW Wicaksono, and DR Akbi, "Interpretable Machine Learning Model For Heart Disease Prediction,"*Procedia Comput Sci*, vol. 227, pp. 439–445, 2023, doi: 10.1016/j.procs.2023.10.544.
[33] M. Harsha Vardhan, M. Rajesh Kumar, M. Vardhini, S. Leela Varalakshmi, and M. Kumar, "Heart Disease Prediction Using Machine Learning," 2023. [Online]. Available: https://jespublication.com/
[34] Y. Huang*et al.*, "Using a machine learning-based risk prediction model to analyse the coronary artery calcification score and predict coronary heart disease and risk assessment," Comput Biol Med, vol. 151, p. 106297, Dec. 2022, doi: 10.1016/j.compbiomed.2022.106297.
[35] G. Manikandan, B. Pragadeesh, V. Manojkumar, AL Karthikeyan, R. Manikandan, and AH Gandomi, "Classification models combined with Boruta feature selection for heart disease prediction," *Inform Med Unlocked*, vol. 44, p. 101442, 2024, doi: 10.1016/j.imu.2023.101442.
[36] Md. A. Islam, Md. ZH Majumder, and Md. A. Hussein, "Chronic kidney disease prediction based on machine learning algorithms," *J Pathol Inform*, vol. 14, p. 100189, 2023, doi: 10.1016/j.jpi.2023.100189.
[37] Md. Mustafizur Rahman, Md. Al-Amin, and J. Hossain, "Machine learning models for chronic kidney disease diagnosis and prediction," *Biomed Signal Process Control*, vol. 87, p. 105368, Jan. 2024, doi: 10.1016/j.bspc.2023.105368.
[38] B. Khan, R. Naseem, F. Muhammad, G. Abbas, and S. Kim, "An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy," *IEEE Access*, vol. 8, pp. 55012–55022, 2020, doi: 10.1109/ACCESS.2020.2981689.
[39] Y. Li, X. Wu, P. Yang, G. Jiang, and Y. Luo, "Machine Learning for Lung Cancer Diagnosis, Treatment, and Prognosis," *Genomics Proteomics Bioinformatics*, vol. 20, no. 5, pp. 850–866, Oct. 2022, doi: 10.1016/j.gpb.2022.11.003.
[40] H. Barnes*et al.*, "Machine learning in radiology: the new frontier in interstitial lung diseases," Lancet Digit Health, vol. 5, no. 1, pp. e41–e50, Jan. 2023, doi: 10.1016/S2589-7500(22)00230-8.
[41] B. Huang*et al.*, "Prediction of lung malignancy progression and survival with machine learning based on pretreatment FDG-PET/CT," EBioMedicine, vol. 82, p. 104127, Aug. 2022, doi: 10.1016/j.ebiom.2022.104127.
[42] FJM Shamrat, S. Azam, A. Karim, K. Ahmed, FM Bui, and F. De Boer, "High-precision multiclass classification of lung disease through customised MobileNetV2 from chest X-ray images,"*Comput Biol Med*, vol. 155, p. 106646, Mar. 2023, doi: 10.1016/j.compbiomed.2023.106646.
[43] S. Rahman, FM Javed, M. Shamrat, Z. Tasnim, J. Roy, and SA Hossain, "A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms," *International Journal Of Scientific & Technology Research*, vol. 8, no. 11, 2019, [Online]. Available: www.ijstr.org
[44] S. Skurzak*et al.*, "A simple machine learning-derived rule to promote ERAS pathways in Liver Transplantation," Journal of Liver Transplantation, vol. 12, p. 100179, Nov. 2023, doi: 10.1016/j.liver.2023.100179.
[45] J. Singh, S. Bagga, and R. Kaur, "Software-based Prediction of Liver Disease with Feature Selection and Classification Techniques," *Procedia Comput Sci*, vol. 167, pp. 1970–1980, 2020, doi: 10.1016/j.procs.2020.03.226.
[46] C. Chakraborty, M. Bhattacharya, S. Pal, and S.-S. Lee, "From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare," *Curr Res Biotechnol*, vol. 7, p. 100164, 2024, doi: 10.1016/j.crbiot.2023.100164.
[47] A. Javaid*et al.*, "Medicine 2032: The future of cardiovascular disease prevention with machine learning and digital health technology," Am J Prev Cardiol, vol. 12, p. 100379, Dec. 2022, doi: 10.1016/j.ajpc.2022.100379.

[48] AM Alaa, T. Bolton, E. Di Angelantonio, J. H. F. Rudd, and M. van der Schaar, "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants," *PLOS One*, vol. 14, no. 5, p. e0213653, May 2019, doi: 10.1371/journal.pone.0213653.
[49] KY Ngiam and IW Khor, "Big data and machine learning algorithms for healthcare delivery," *Lancet Oncol*, vol. 20, no. 5, pp. e262–e273, May 2019, doi: 10.1016/S1470-2045(19)30149-4.
[50] A. Alanazi, "Using machine learning for healthcare challenges and opportunities," *Inform Med Unlocked*, vol. 30, p. 100924, 2022, doi: 10.1016/j.imu.2022.100924.
[51] K. Rasheed, A. Qayyum, M. Ghaly, A. Al-Fuqaha, A. Razi, and J. Qadir, "Explainable, trustworthy, and ethical machine learning for healthcare: A survey," *Comput Biol Med*, vol. 149, p. 106043, Oct. 2022, doi: 10.1016/j.compbiomed.2022.106043.

BIOGRAPHY OF AUTHORS

	**Rosyid Ridlo Al-Hakim,** as a graduate student at Primate Research Center, IPB University (PSSP-IPB). Research interest in applied artificial intelligence, wildlife technology/ecological informatics, forensics, and information systems. Email: alhakimrosyid@apps.ipb.ac.id, Orcid ID: 0000-0003-1502-4745.
	**Yurii Prokopchuk,** PhD, Doctor of Technical Sciences Professor, Pridneprovsk State Academy of Civil Engineering and Architecture, Department of Computer Science, Information Technologies and Applied Mathematics Chernyshevsky street, 24a, Dnipro, Ukraine, 49600 E-mail: itk3@ukr.net ORCID: http://orcid.org/0000-0002-8544-1838 ResearchGate: https://www.researchgate.net/profile/Yurii-Prokopchuk.