Below is a summary table compiling the key points from the literature review file. Each row represents one research paper.

Title & Authors	Proposed System	Algorithms Used	Methodology for Increasing Accuracy	Accuracy Generated	Limitations	Future Scope
Disease Prediction using Machine Learning(Kriti Gandhi, Mansi Mittal, Neha Gupta, Shafali Dhall, June 2020)	A classification-based ML system that trains multiple models on healthcare data to predict diseases from symptoms and medical history, emphasizing early diagnosis and effective treatment.	K-Nearest Neighbors (KNN), Logistic Regression, Decision Tree, Naïve Bayes, Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Random Forest.	Utilizes feature selection techniques (Recursive Feature Elimination and embedded methods), data preprocessing (handling missing values, train/test split) to refine input data.	Highest accuracy achieved by Logistic Regression at 98.87%; Random Forest performed worst at 80.85%.	Risk of overfitting; limited dataset (133 columns, 40 diseases); real-world data may be noisy and incomplete.	Integration of deep learning for enhanced feature extraction, expansion with real-time patient data, and incorporating wearable health monitoring systems.
Disease Prediction using Machine Learning(Palle Pramod Reddy, Dirisinala Madhu	An automated software solution focused on chronic disease prediction using both	Primarily Random Forest Classifier (with data preprocessing steps such as forward fill,	Thorough data preprocessing (null handling, data standardization) and employing	Reported accuracies vary by disease: Diabetes & Breast Cancer at 98.25%, Heart	Dependence on online data sources may affect accuracy; challenges with processing unstructure	Future improvements include enhancing model accuracy (especially for diseases with lower

Title & Authors	Proposed System	Algorithms Used	Methodology for Increasing Accuracy	Accuracy Generated	Limitations	Future Scope
Babu, Hardeep Kumar, Dr. Shivi Sharma, May 2021)	structured and unstructured data, including a latent factor model to handle missing values and consultation with experts to select features.	standardization, and splitting into training/testing sets).	Random Forest for both feature selection and prediction from diverse online data sources.	Disease at 85.25%, Kidney Disease at 99%, Liver Disease at 78%.	d text; potential issues with generalizability across regions.	accuracy), expanding the range of diseases covered, and exploring hybrid models for better performance.
Human Disease Prediction using Machine Learning Techniques and Real-life Parameters (K. Gaurav, A. Kumar, P. Singh, A. Kumari, M. Kasar, T. Suryawanshi, June 2023)	A system that predicts human diseases by leveraging real-life parameters—including symptoms, demographics, and lifestyle factors—by integrating structured and unstructured data	Random Forest, Long Short-Term Memory (LSTM), and Support Vector Machine (SVM).	Implements hyperparameter tuning (especially for Random Forest), assigns weighted values to rare symptoms based on geographic distribution, and uses LSTM for time-series analysis of patient history	Random Forest achieved a highest accuracy of 97% (with other models like Weighted KNN, Naïve Bayes, and SVM scoring 93.5%, 94.8%, and 90% respectively).	Heavily relies on structured datasets; higher computational cost; model accuracy can be affected by missing or inaccurate patient data.	Future work focuses on incorporating real-time electronic health records (EHRs), integrating advanced deep learning models, and improving model interpretability for enhanced clinical

Title & Authors	Proposed System	Algorithms Used	Methodology for Increasing Accuracy	Accuracy Generated	Limitations	Future Scope
	sources to aid early diagnosis and reduce clinical workload.		alongside standard feature selection.			decision-making.

Source: Litterature Review File.pdf □cite□turn0file0□