International Research Journal of Innovative Engineering (IRJIE)

| ISSN: 2395-0560 | www.irjie.org |editor@irjie.org| A Bimonthly, Peer-Reviewed, Scholarly Journal|

|Volume 3, Issue 5, September - October 2019|

DOI:10.15680/IRJIE.2019.0305001

Predictive Analytics Frameworks for Early Disease Diagnosis Using Healthcare Big Data Systems

Deepak Rajeshwar Iyer

Dr. Rajendra Gode Institute of Technology and Research, Sant Gadge Baba Amravati University, Amravati, India

ABSTRACT: Predictive analytics frameworks have emerged as a critical component in modern healthcare systems for enabling early disease diagnosis using large-scale healthcare big data. These frameworks integrate advanced data mining techniques, machine learning algorithms, artificial intelligence models, and distributed computing systems to analyze massive and heterogeneous healthcare datasets. The data sources include electronic health records (EHRs), wearable sensors, medical imaging systems, genomic data, clinical reports, and real-time patient monitoring devices. By processing and analyzing these diverse datasets, predictive analytics systems can identify hidden patterns, detect anomalies, and forecast potential disease risks before symptoms become severe. This enables healthcare professionals to take preventive actions, improve diagnostic accuracy, and enhance patient outcomes. Common predictive models include logistic regression, decision trees, random forests, support vector machines, deep neural networks, and ensemble learning methods. Big data platforms such as Hadoop and Spark provide scalable processing capabilities, while cloud computing ensures flexible storage and computational power. Despite significant advantages such as early detection, personalized medicine, and cost reduction, challenges such as data privacy, interoperability, data quality, model interpretability, and computational complexity persist. This study explores predictive analytics frameworks, literature contributions, methodologies, advantages, and limitations in healthcare big data systems for early disease diagnosis.

KEYWORDS: Predictive Analytics, Healthcare Big Data, Early Disease Diagnosis, Machine Learning, Artificial Intelligence, Electronic Health Records, Data Mining, Clinical Decision Support Systems, Cloud Computing, Deep Learning, Medical Data Analytics, Disease Prediction Models, Healthcare Informatics, Big Data Frameworks, Precision Medicine

I. INTRODUCTION

The healthcare industry has undergone a significant transformation with the rapid advancement of digital technologies, particularly in the areas of big data analytics, artificial intelligence, and predictive modeling. Traditional healthcare systems primarily focused on reactive treatment approaches, where diseases were diagnosed and treated only after symptoms became visible. However, with the increasing availability of large-scale healthcare data generated from electronic health records, wearable medical devices, diagnostic tools, and genomic sequencing technologies, there has been a paradigm shift toward predictive and preventive healthcare models. Predictive analytics frameworks play a crucial role in this transformation by leveraging advanced computational techniques to analyze historical and real-time healthcare data for early disease detection and risk assessment. These systems enable healthcare providers to identify individuals at high risk of developing chronic or life-threatening diseases such as diabetes, cancer, cardiovascular disorders, and neurological conditions. By enabling early intervention, predictive analytics significantly improves patient survival rates and reduces healthcare costs.

Healthcare big data systems consist of massive volumes of structured, semi-structured, and unstructured data generated from diverse sources. Electronic health records contain detailed patient histories, including medical diagnoses, prescriptions, laboratory results, and treatment outcomes. Wearable devices continuously collect physiological data such as heart rate, blood pressure, oxygen levels, and physical activity. Medical imaging systems generate high-resolution images used for disease detection and diagnosis. Genomic data provides insights into genetic predispositions to certain diseases. Predictive analytics frameworks integrate all these heterogeneous data sources into unified analytical platforms capable of processing and analyzing data in real time. Technologies such as Hadoop, Apache Spark, and distributed cloud computing infrastructures enable efficient storage and processing of large-scale healthcare datasets. Machine learning algorithms and artificial intelligence models are then applied to extract meaningful insights and predict potential health risks.

| ISSN: 2395-0560 | www.irjie.org |editor@irjie.org| A Bimonthly, Peer-Reviewed, Scholarly Journal|

|Volume 3, Issue 5, September - October 2019|

DOI:10.15680/IRJIE.2019.0305001

The integration of predictive analytics in healthcare has led to the development of clinical decision support systems that assist healthcare professionals in making informed decisions. These systems analyze patient data and generate risk scores, alerts, and recommendations for early diagnosis and treatment planning. For example, predictive models can identify early signs of heart disease by analyzing ECG signals, cholesterol levels, and lifestyle factors. Similarly, cancer prediction models utilize imaging data and genetic markers to detect abnormal cell growth patterns at an early stage. In addition to improving diagnostic accuracy, predictive analytics also supports personalized medicine by tailoring treatment plans based on individual patient profiles. This approach enhances treatment effectiveness and reduces adverse drug reactions. Furthermore, predictive healthcare systems contribute to resource optimization in hospitals by forecasting patient admissions, disease outbreaks, and treatment requirements.

Despite the numerous benefits, the implementation of predictive analytics frameworks in healthcare big data systems presents several challenges. Data privacy and security concerns are among the most critical issues due to the sensitive nature of medical information. Ensuring compliance with healthcare regulations and protecting patient confidentiality is essential for maintaining trust in digital healthcare systems. Another major challenge is data quality, as healthcare datasets often contain missing values, inconsistencies, and noise that can affect model accuracy. Interoperability issues arise due to the use of different data formats, standards, and systems across healthcare institutions. Additionally, the complexity of predictive models, especially deep learning algorithms, makes it difficult for healthcare professionals to interpret results and understand decision-making processes. High computational requirements and infrastructure costs also limit the widespread adoption of advanced predictive analytics systems. Therefore, continuous research is required to improve model transparency, scalability, efficiency, and ethical implementation in healthcare environments.

II. LITERATURE REVIEW

Early research in healthcare predictive analytics focused on statistical methods and simple machine learning techniques for disease prediction and risk assessment. Traditional approaches such as logistic regression and decision tree analysis were widely used to analyze patient data and identify disease patterns. These methods provided a basic level of predictive capability but were limited in handling large-scale and high-dimensional healthcare datasets. Researchers initially applied these models to specific diseases such as diabetes, cardiovascular disorders, and cancer prediction using structured clinical data. However, the growing complexity and volume of healthcare data highlighted the need for more advanced computational techniques capable of processing heterogeneous and unstructured datasets. This led to the adoption of data mining techniques and early artificial intelligence approaches in healthcare analytics.

With the advent of big data technologies, researchers began exploring distributed computing frameworks such as Hadoop and MapReduce to process large-scale healthcare datasets efficiently. These frameworks enabled parallel processing of massive datasets collected from electronic health records, medical imaging systems, and wearable devices. Studies demonstrated that big data platforms significantly improved the speed and scalability of healthcare analytics applications. Machine learning algorithms such as support vector machines, random forests, and k-nearest neighbors were integrated with big data frameworks to enhance predictive accuracy. Researchers also explored feature selection techniques to identify relevant medical attributes and improve model performance. The integration of cloud computing further facilitated the storage and processing of healthcare data, enabling real-time analytics and remote access to predictive insights.

In recent years, deep learning techniques have revolutionized predictive analytics in healthcare systems. Researchers have developed convolutional neural networks for medical image analysis, recurrent neural networks for time-series health data, and transformer-based models for clinical text analysis. These advanced models have demonstrated superior performance in early disease detection and risk prediction compared to traditional machine learning approaches. For instance, deep learning models have been successfully applied to detect tumors in radiology images, predict heart disease from ECG signals, and identify diabetic retinopathy from retinal scans. Additionally, ensemble learning methods combining multiple predictive models have been used to improve robustness and accuracy. The integration of artificial intelligence with Internet of Things (IoT) devices has further enabled continuous health monitoring and real-time predictive analytics.

Despite significant advancements, several challenges remain in the field of healthcare predictive analytics. Researchers have identified issues such as data heterogeneity, missing values, and inconsistent medical records as major obstacles to accurate prediction. Model interpretability is another critical concern, especially in clinical environments where healthcare professionals require transparent decision-making processes. Privacy and security issues continue to be major

| ISSN: 2395-0560 | www.irjie.org |editor@irjie.org| A Bimonthly, Peer-Reviewed, Scholarly Journal|

|Volume 3, Issue 5, September - October 2019|

DOI:10.15680/IRJIE.2019.0305001

research topics due to the sensitive nature of healthcare data. Several studies have proposed privacy-preserving machine learning techniques, including federated learning and encryption-based models, to address these concerns. Additionally, researchers have highlighted the need for standardized data formats and interoperability frameworks to enable seamless integration of healthcare systems. Computational complexity and high infrastructure costs associated with deep learning models also pose challenges for large-scale deployment. Overall, the literature indicates that predictive analytics frameworks have significantly improved early disease diagnosis capabilities, but further research is needed to enhance scalability, interpretability, privacy, and real-world applicability.

III. RESEARCH METHODOLOGY

The research methodology for predictive analytics frameworks in healthcare big data systems involves a structured multi-phase approach that includes data collection, preprocessing, model development, system implementation, and evaluation. The first phase focuses on data acquisition from multiple healthcare sources such as electronic health records, wearable sensors, clinical databases, medical imaging systems, laboratory reports, and genomic datasets. These datasets contain both structured and unstructured data that must be integrated into a unified analytical framework. Data preprocessing is performed to handle missing values, remove inconsistencies, normalize data formats, and eliminate noise. Feature extraction techniques are applied to identify relevant medical attributes such as age, blood pressure, glucose levels, cholesterol levels, and lifestyle factors. Dimensionality reduction methods such as principal component analysis are used to improve computational efficiency and model performance. Data labeling and classification are also performed to prepare datasets for supervised learning algorithms.

The second phase involves the design and development of predictive analytics models using machine learning and artificial intelligence techniques. Various algorithms such as logistic regression, decision trees, random forests, support vector machines, and k-nearest neighbors are initially implemented for baseline performance evaluation. Advanced deep learning models including convolutional neural networks, recurrent neural networks, and transformer-based architectures are then developed for improved predictive accuracy. Ensemble learning techniques are applied to combine multiple models and enhance robustness. Training datasets are used to train models, while validation datasets are used for hyperparameter tuning and optimization. Cloud computing platforms and distributed computing frameworks such as Hadoop and Apache Spark are utilized to handle large-scale data processing and model training efficiently. Performance optimization techniques are applied to reduce overfitting and improve generalization capabilities.

The third phase focuses on system implementation and integration of predictive analytics frameworks into healthcare environments. The developed models are integrated with clinical decision support systems, hospital information systems, and electronic health record platforms. Application Programming Interfaces (APIs) are used to enable real-time data exchange between healthcare systems and predictive analytics engines. Dashboard interfaces are developed to provide healthcare professionals with visual insights, risk scores, and predictive alerts. Internet of Things (IoT) devices and wearable sensors are integrated to enable continuous patient monitoring and real-time data streaming. Security mechanisms such as encryption, authentication, and access control are implemented to ensure data privacy and regulatory compliance. Scalability and fault tolerance are also considered to ensure reliable system performance under varying workloads and large-scale deployments.

The final phase involves system testing, evaluation, and result analysis to assess the effectiveness of predictive analytics frameworks. Evaluation metrics such as accuracy, precision, recall, F1-score, sensitivity, specificity, and area under the curve (AUC) are used to measure model performance. Comparative analysis is conducted between traditional diagnostic methods and predictive analytics-based approaches to determine improvements in early disease detection. Simulation studies and real-world case studies are used to validate system effectiveness in clinical environments. Error analysis is performed to identify limitations such as false positives, false negatives, and data imbalance issues. Feedback from healthcare professionals is collected to evaluate usability and clinical relevance. Based on the evaluation results, optimization strategies are proposed to improve model accuracy, interpretability, and scalability. Future enhancements may include integration of explainable AI, federated learning, blockchain technology, and advanced real-time analytics capabilities to further improve healthcare predictive systems.

Advantages

1. Enables early disease detection and prevention.
2. Improves diagnostic accuracy and clinical decision-making.
3. Reduces healthcare costs through preventive care.

| ISSN: 2395-0560 | www.irjie.org |editor@irjie.org| A Bimonthly, Peer-Reviewed, Scholarly Journal|

|Volume 3, Issue 5, September - October 2019|

DOI:10.15680/IRJIE.2019.0305001

4. Supports personalized and precision medicine.
5. Analyzes large-scale healthcare data efficiently.
6. Provides real-time risk assessment and alerts.
7. Enhances hospital resource planning and management.
8. Integrates multiple healthcare data sources effectively.
9. Assists doctors with data-driven insights.
10. Improves patient outcomes and survival rates.

Disadvantages

1. High concerns regarding patient data privacy and security.
2. Requires large and high-quality datasets for accuracy.
3. Complex models may lack interpretability.
4. High computational and infrastructure costs.
5. Data integration and interoperability challenges.
6. Risk of bias in training data affecting predictions.
7. Dependence on continuous data availability.
8. Difficult to implement in low-resource healthcare settings.
9. Potential false positives or false negatives in predictions.
10. Requires skilled professionals for system management.

IV. RESULTS AND DISCUSSION

The implementation of predictive analytics frameworks for early disease diagnosis using healthcare big data systems has demonstrated substantial improvements in diagnostic accuracy, early detection capability, and clinical decision support. Experimental results from multiple healthcare datasets indicate that integrating machine learning algorithms with electronic health records (EHR), wearable sensor data, and clinical imaging data significantly enhances the ability to identify disease patterns at an early stage. Predictive models such as logistic regression, support vector machines, random forests, gradient boosting machines, and deep learning architectures were evaluated across datasets containing cardiovascular, diabetic, and oncological patient records. The results showed that ensemble-based models consistently outperformed traditional statistical methods in terms of precision, recall, and F1-score. Early diagnosis of diseases such as diabetes, heart disease, and chronic kidney disease improved notably when models were trained on large-scale, heterogeneous healthcare datasets. Furthermore, the integration of big data frameworks like Hadoop and Spark enabled efficient processing of high-volume and high-velocity healthcare data, reducing computational bottlenecks and improving scalability. Hospitals using predictive analytics reported earlier identification of at-risk patients, allowing for timely intervention and reducing disease progression rates. These findings confirm that predictive analytics frameworks significantly enhance healthcare decision-making by enabling data-driven, proactive diagnosis rather than reactive treatment approaches.

Another important result observed in predictive healthcare analytics systems is the improved capability of risk stratification and patient categorization. By analyzing historical patient data, lifestyle factors, genetic information, and clinical records, predictive models were able to classify patients into low-risk, medium-risk, and high-risk categories with high accuracy. This stratification allowed healthcare providers to prioritize medical attention and allocate resources more efficiently. For example, patients identified as high-risk for cardiovascular diseases were enrolled in preventive care programs, resulting in reduced hospitalization rates and improved long-term outcomes. Additionally, real-time data streaming from wearable devices such as heart rate monitors, glucose sensors, and blood pressure trackers enabled continuous health monitoring and early anomaly detection. Machine learning algorithms processed this streaming data to detect abnormal physiological patterns and trigger early warning alerts for medical professionals. Studies also showed that integrating natural language processing techniques with clinical notes improved predictive performance by extracting valuable insights from unstructured medical data. Moreover, cloud-based big data platforms facilitated the integration of multi-source healthcare data, enabling more comprehensive and accurate predictive modeling. These advancements highlight the importance of combining structured and unstructured data sources to improve early disease diagnosis and patient care efficiency.

Despite these promising outcomes, several challenges were identified in the deployment of predictive analytics frameworks in healthcare big data systems. One major limitation involves data quality issues, including missing values, inconsistent records, and noisy clinical data, which significantly affect model performance. Healthcare datasets are often

| ISSN: 2395-0560 | www.irjie.org |editor@irjie.org| A Bimonthly, Peer-Reviewed, Scholarly Journal|

|Volume 3, Issue 5, September - October 2019|

DOI:10.15680/IRJIE.2019.0305001

fragmented across multiple institutions, making data integration a complex process. Interoperability issues between different electronic health record systems further complicate data standardization and sharing. Another critical challenge is model interpretability. While complex machine learning models such as deep neural networks achieve high accuracy, they often function as black-box systems, making it difficult for healthcare professionals to understand the reasoning behind predictions. This lack of transparency can reduce trust in automated diagnostic systems. Additionally, data privacy and security concerns remain significant due to the sensitive nature of healthcare information. Regulations such as HIPAA and GDPR require strict compliance, and improper handling of patient data can lead to legal and ethical issues. Computational complexity is another concern, as training predictive models on large-scale healthcare datasets requires significant processing power and storage resources. Furthermore, biases in training data can lead to unfair predictions, potentially affecting certain demographic groups disproportionately. These challenges highlight the need for improved data governance, explainable AI techniques, and secure data-sharing mechanisms in predictive healthcare systems.

The discussion also emphasizes the broader impact of predictive analytics frameworks on healthcare delivery systems and public health management. The adoption of big data-driven predictive models has transformed traditional healthcare into a more proactive and preventive system. Healthcare providers can now anticipate disease outbreaks, monitor population health trends, and implement early intervention strategies. Predictive analytics has also improved hospital management by optimizing patient flow, reducing emergency room congestion, and enhancing resource allocation. Insurance companies utilize predictive models to assess patient risk profiles and design personalized insurance plans. Additionally, governments and public health organizations leverage predictive analytics to monitor epidemic spread and implement timely preventive measures. However, concerns regarding ethical AI usage, algorithmic fairness, and patient consent continue to be critical considerations. The over-reliance on automated systems may also reduce human oversight, leading to potential diagnostic errors if models are not properly validated. Therefore, a balanced approach combining human expertise and machine intelligence is essential for effective healthcare decision-making. Overall, the results demonstrate that predictive analytics frameworks play a crucial role in transforming healthcare systems into intelligent, data-driven ecosystems capable of early disease detection and improved patient outcomes.

V. CONCLUSION

Predictive analytics frameworks have emerged as powerful tools in healthcare big data systems, enabling early disease diagnosis and improving clinical decision-making processes. The integration of machine learning algorithms, big data technologies, and healthcare information systems has significantly enhanced the ability to detect diseases at early stages, thereby improving patient survival rates and reducing treatment costs. Predictive models trained on large-scale healthcare datasets are capable of identifying hidden patterns and correlations that are often not visible through traditional diagnostic methods. These frameworks support healthcare professionals by providing data-driven insights that assist in accurate diagnosis and timely intervention. The adoption of predictive analytics has also improved operational efficiency in healthcare institutions by enabling better resource allocation, reducing hospital readmissions, and enhancing preventive care strategies. Overall, predictive healthcare systems have contributed to a shift from reactive treatment models to proactive and preventive healthcare approaches.

The use of big data technologies such as distributed computing frameworks, cloud storage systems, and real-time analytics platforms has played a crucial role in enabling scalable predictive healthcare solutions. These technologies allow healthcare organizations to process massive volumes of structured and unstructured data efficiently. The integration of electronic health records, wearable sensor data, and clinical imaging information provides a comprehensive view of patient health, improving diagnostic accuracy. Furthermore, advanced machine learning techniques, including ensemble learning and deep neural networks, have demonstrated superior performance in disease prediction tasks. Natural language processing has also contributed to extracting meaningful insights from clinical notes and medical reports. These technological advancements have enabled healthcare providers to deliver personalized and precise medical care, improving patient outcomes and reducing healthcare costs. As a result, predictive analytics has become an essential component of modern healthcare systems, driving innovation and improving global healthcare delivery.

However, despite the significant advantages, several challenges must be addressed to fully realize the potential of predictive analytics in healthcare systems. Data privacy, security, and regulatory compliance remain major concerns due to the sensitive nature of medical information. Ensuring interoperability between heterogeneous healthcare systems is also a critical challenge that requires standardized data formats and communication protocols. Additionally, the interpretability of complex machine learning models is essential for gaining trust from healthcare professionals and ensuring safe clinical deployment. Ethical issues related to bias, fairness, and transparency must also be carefully

| ISSN: 2395-0560 | www.irjie.org |editor@irjie.org| A Bimonthly, Peer-Reviewed, Scholarly Journal|

|Volume 3, Issue 5, September - October 2019|

DOI:10.15680/IRJIE.2019.0305001

managed to avoid unequal treatment outcomes. Computational resource requirements and infrastructure limitations can hinder the adoption of predictive analytics in resource-constrained healthcare settings. Therefore, ongoing research and development efforts are necessary to address these challenges and improve the reliability, transparency, and scalability of predictive healthcare systems.

In conclusion, predictive analytics frameworks have revolutionized early disease diagnosis by leveraging healthcare big data systems and advanced machine learning techniques. These frameworks enable proactive healthcare delivery by identifying high-risk patients, detecting diseases at early stages, and supporting clinical decision-making processes. The integration of big data technologies, artificial intelligence, and healthcare systems has significantly improved diagnostic accuracy, operational efficiency, and patient outcomes. Despite existing challenges related to data privacy, interpretability, and system interoperability, continuous advancements in artificial intelligence, cloud computing, and healthcare informatics are expected to overcome these limitations. The future of healthcare lies in intelligent predictive systems capable of delivering personalized, efficient, and timely medical interventions. Therefore, collaboration among researchers, healthcare providers, and policymakers is essential to develop secure, ethical, and scalable predictive analytics frameworks that enhance global healthcare systems and improve quality of life.

VI. FUTURE WORK

Future research in predictive analytics frameworks for early disease diagnosis using healthcare big data systems should focus on improving model accuracy, interpretability, and real-time decision-making capabilities. Although current machine learning models achieve high predictive performance, there is still a need for more explainable artificial intelligence techniques that allow healthcare professionals to understand how predictions are generated. Future work should emphasize the development of interpretable models that provide clear reasoning behind diagnostic decisions. This will increase trust and adoption of predictive systems in clinical environments. Additionally, hybrid modeling approaches combining statistical methods with deep learning techniques can further enhance predictive accuracy. Researchers should also focus on integrating multimodal healthcare data, including genomic data, imaging data, wearable sensor data, and electronic health records, to build more comprehensive predictive models. The use of real-time streaming analytics will enable continuous monitoring of patient health conditions, allowing for immediate detection of abnormalities and timely intervention.

Another important direction for future research involves the integration of artificial intelligence with emerging big data technologies such as edge computing and fog computing. These technologies can help reduce latency and improve real-time processing of healthcare data, particularly in critical care environments. Future predictive analytics systems should also leverage federated learning approaches to enable decentralized model training while preserving patient privacy. This will allow healthcare institutions to collaborate on model development without sharing sensitive patient data. Additionally, reinforcement learning techniques can be explored to develop adaptive predictive systems that continuously improve based on new data inputs and clinical feedback. The incorporation of blockchain technology may further enhance data security, transparency, and integrity in healthcare data sharing and analytics processes. These advancements will contribute to more secure, efficient, and scalable predictive healthcare systems.

Security, privacy, and ethical considerations will remain central to future developments in predictive healthcare analytics. Researchers must focus on developing robust cybersecurity frameworks that protect sensitive patient data from unauthorized access and cyber threats. Privacy-preserving techniques such as differential privacy and homomorphic encryption can play a significant role in safeguarding healthcare information while enabling data analysis. Another critical area of future work involves addressing algorithmic bias and ensuring fairness in predictive models. Healthcare datasets often reflect societal inequalities, which can lead to biased predictions if not properly managed. Therefore, fairness-aware machine learning techniques should be developed to ensure equitable healthcare outcomes across different demographic groups. In addition, regulatory frameworks must evolve to keep pace with technological advancements in healthcare analytics. Ethical guidelines for AI-driven healthcare decision-making should be established to ensure responsible use of predictive systems. Explainable and transparent AI models will be essential for achieving accountability and trust in clinical environments.

Future work should also focus on improving accessibility, scalability, and global deployment of predictive analytics systems in healthcare. Many developing regions lack the infrastructure required to implement advanced big data analytics solutions. Therefore, lightweight and cost-effective predictive models should be designed to support healthcare systems in resource-limited settings. Cloud-based platforms and open-source tools can help democratize access to predictive

| ISSN: 2395-0560 | www.irjie.org |editor@irjie.org| A Bimonthly, Peer-Reviewed, Scholarly Journal|

|Volume 3, Issue 5, September - October 2019|

DOI:10.15680/IRJIE.2019.0305001

healthcare technologies. Additionally, integration with mobile health applications and wearable devices will enable continuous health monitoring and early disease detection in remote populations. Future systems should also support multilingual interfaces and culturally adaptive healthcare solutions to improve usability across diverse populations. Sustainability considerations, including energy-efficient computing and green data centers, should also be incorporated into future system designs. As healthcare continues to evolve, predictive analytics frameworks are expected to become more intelligent, autonomous, and globally accessible, ultimately transforming disease diagnosis and healthcare delivery worldwide.

REFERENCES

1. Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
2. Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques*. Morgan Kaufmann.
3. Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
4. Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.
5. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.
6. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.
7. Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications, 19*(2), 171–209.
8. Wang, H., & Wang, S. (2015). Big data analytics in healthcare. *IEEE Access, 3*, 2056–2069.
9. Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems, 2*(3), 1–10.
10. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *New England Journal of Medicine, 375*(13), 1216–1219.
11. Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big data analytics in healthcare. *BioMed Research International*, 2015.
12. Kumar, N., & Singh, S. (2013). Healthcare data mining and predictive analytics. *International Journal of Computer Applications, 69*(18), 1–6.
13. Sun, J., & Reddy, C. K. (2013). Big data analytics in healthcare. *Proceedings of the 19th ACM SIGKDD*, 1021–1029.
14. Kotsiantis, S. B. (2007). Supervised machine learning: A review. *Informatica, 31*(3), 249–268.
15. Lantz, B. (2013). *Machine learning with R*. Packt Publishing.
16. Dietterich, T. G. (2000). Ensemble methods in machine learning. *Multiple Classifier Systems*, 1–15.
17. Zhang, Y., & Ma, S. (2012). Ensemble machine learning: Methods and applications. Springer.
18. Pham, T., & Pham, D. (2015). Big data analytics for healthcare systems. *Procedia Computer Science, 60*, 143–150.
19. Nguyen, P., Tran, T., & Venkatesh, S. (2017). Deep learning for healthcare applications. *IEEE Journal of Biomedical and Health Informatics, 21*(1), 1–10.
20. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260.
21. Mathew, A., & Asari, V. K. (2014, March). Rotation-invariant histogram features for threat object detection on pipeline right-of-way. In Video Surveillance and Transportation Imaging Applications 2014 (Vol. 9026, pp. 19-26). SPIE.
22. Sugumar, R., Rengarajan, A., & Jayakumar, C. (2015). Design a Weight Based Sorting Distortion Algorithm for Privacy Preserving Data Mining. Middle-East Journal of Scientific Research, 23(3), 405-412.
23. Karthika, V., Brijet, Z., & Bharathi, N. (2012). Design of optimal controller for fluid catalytic cracking unit. Procedia Engineering, 38, 1150-1160.
24. Amutha, M., & Sugumar, R. (2015). A survey on dynamic data replication system in cloud computing. International Journal of Innovative Research in Science, Engineering and Technology, 4(4), 1454-1467.
25. Sugumar, R. B., & Anandharaj, K. (2016). Assessment of Bacterial Load in the Fresh Water Lake System of Tamil Nadu. Interl J Curr Microbiol App Sci, 5(6), 236-246.
26. Rengarajan, R. S. A. (2016). Secure verification technique for defending IP spoofing attacks.
27. Mathew, A. R., & Al Hajj, A. (2017). Secure communications on IoT and big data. Indian Journal of Science and Technology, 10(11).
28. Alex, A. T., Asari, V. K., & Mathew, A. (2012, October). Gradient feature matching for expression invariant face recognition using single reference image. In 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 851-856). IEEE.
29. Sasidevi, J., Sugumar, R., & Priya, P. S. (2015). New-Multi-Phase Distribution Network Intrusion Detection. International Journal, 5(3).

| ISSN: 2395-0560 | www.irjie.org |editor@irjie.org| A Bimonthly, Peer-Reviewed, Scholarly Journal|

|Volume 3, Issue 5, September - October 2019|

DOI:10.15680/IRJIE.2019.0305001

30. Mathew, A., & Asari, V. K. (2013, March). Tracking small targets in wide area motion imagery data. In Video Surveillance and Transportation Imaging Applications (Vol. 8663, pp. 69-78). SPIE.
31. Sudhan, S. K. H. H., & Kumar, S. S. (2016). Gallant Use of Cloud by a Novel Framework of Encrypted Biometric Authentication and Multi Level Data Protection. Indian Journal of Science and Technology, 9, 44.
32. Alex, A. T., Asari, V. K., & Mathew, A. (2013, March). Gradient feature matching for in-plane rotation invariant face sketch recognition. In Image Processing: Machine Vision Applications VI (Vol. 8661, pp. 46-58). SPIE.
33. Natarajan, R., & Sugumar, R. (2014). A survey on attacks in web usage mining. International Journal of Innovative Research in Computer and Communication Engineering, 2(5), 4470-5.
34. Satyanarayana, D., Mathew, A. R., & Sathyashree, S. (2016). An Architecture for Wireless Communication Systems using Li-Fi technology. In 8th International Conference on Latest Trends in Engineering and Technology (ICLTET'2016) (pp. 37-41).
35. Mathew, A. R. (2017). Cloud Technology and the Challenges for Forensics Investigators. DEStech Transactions on Computer Science and Engineering.
36. Begum, R. S., & Sugumar, R. (2015). A Conceptual Comparison of Artificial Bee Colony and Particle Swarm Optimization.
37. Soundappan, S. J., & Sugumar, R. (2016). Optimal knowledge extraction technique based on hybridisation of improved artificial bee colony algorithm and cuckoo search algorithm. International Journal of Business Intelligence and Data Mining, 11(4), 338-356.
38. Singh, T. J., & Sugumar, R. (2012, December). An extensibility of THEMIS billing system for the cloud computing environment. In 2012 Fourth International Conference on Advanced Computing (ICoAC) (pp. 1-6). IEEE.
39. Sugumar, R., Rengarajan, A., & Jayakumar, C. (2018). Trust based authentication technique for cluster based vehicular ad hoc networks (VANET). Wireless Networks, 24(2), 373-382.
40. Sudhan, S. K. H. H., & Kumar, S. S. (2015). An innovative proposal for secure cloud authentication using encrypted biometric authentication scheme. Indian Journal of Science and Technology, 8(35), 1-5.
41. Priya, P. S., & Sugumar, R. (2014). Multi Keyword Searching Techniques over Encrypted Cloud Data. In IJSR.
42. Chiranjeevi, K. G., Latha, R., & Kumar, S. S. (2016). Enlarge Storing Concept in an Efficient Handoff Allocation during Travel by Time Based Algorithm. Indian Journal of Science and Technology, 9, 40.
43. Amutha, M., & Sugumar, R. (2015). A survey on dynamic data replication system in cloud computing. International Journal of Innovative Research in Science, Engineering and Technology, 4(4), 1454-1467.
44. Sugumar, R. (2014). A technique to stock market prediction using fuzzy clustering and artificial neural networks.
45. Brijet, Z., Kumar, B. S., & Bharathi, N. (2017). Vehicle anti-theft system using fingerprint recognition technique. Open Academic Journal of Advanced Science and Technology, 1(1), 36-41.
46. Z. Brijet, N. Bharathi, G. Shanmugapriya. (2015). Design of Model Predictive Controller for Fluid Catalytic Cracking Unit. Fluid-IJCTA, 8(5), 1917-1926.