Email: editor@ijerst.com or editor.ijerst@gmail.com

ISSN 2319-5991 www.ijerst.com

Vol. 16, Issue 3, 2020

Optimizing Healthcare Data Streams Using Real-Time Big Data Analytics and AI Techniques

Surendar Rama Sitaraman

Intel Corporation, Folsom, California, USA

sramasitaraman@gmail.com

ABSTRACT

Healthcare delivery gets revolutionized by the integration of Artificial Intelligence (AI) and Big Data Analytics into mobile health (m-Health) technologies. The present investigation looks at the way AI and real-time analytics might improve the administration of healthcare data, focusing especially on neural networks—which processed complex medical data with an astounding 92% accuracy. Rapid data processing was made possible by the combination of Apache Spark and Hadoop, and it's essential for prompt healthcare interventions. Although these developments are encouraging, problems in handling unstructured data from wearable technology and protecting data privacy still exist. The investigation points out areas that require more research and development while also highlighting the potential for AI and Big Data to revolutionize healthcare.

Keywords: m-Health, Big Data Analytics, Artificial Intelligence, Neural Networks, Apache Spark, Hadoop.

1 INTRODUCTION

Mobile health, or m-Health, is a ground-breaking method of providing healthcare in recent years. Through the use of commonplace gadgets such as wearable monitors, cellphones, and personal digital assistants (PDAs), m-Health seeks to improve the efficiency and accessibility of health monitoring and management. The potential of m-Health has been further enhanced by the integration of cutting-edge technologies like artificial intelligence (AI) and big data analytics, which present creative answers to persistent problems in healthcare. The proliferation of health information and new issues an unparalleled increase in health-related data has occurred from the widespread use of m-Health applications. Every encounter adds to a large and varied pool of data, from monitoring daily activities to capturing intricate medical imaging. But because so much of this data is available in unstructured and fragmented formats, it can be challenging for healthcare professionals to draw useful conclusions from it. To fully utilize this data flood, managing and understanding it is now a major challenge that calls for advanced tools and techniques.

Using big data analytics and AI in m-Health big data analytics and artificial intelligence are now effective friends in tackling the complexity of contemporary healthcare data. Massive datasets can be quickly processed and analyzed by AI algorithms, that can then quickly spot patterns and trends that the human eye might miss. These technologies allow for more precise diagnosis, customized treatment regimens, and proactive health management if paired with Big Data frameworks. This investigation explores the many AI methods and Big Data tactics that are currently being used in m-Health systems. inevitably are able to comprehend how these tools are revolutionizing patient care, optimizing clinical workflows, and promoting improved health outcomes by looking at case studies and real-world implementations.

Developments in mobile health technologies a variety of cutting-edge technologies that are intended to monitor and enhance health in real-time are at the core of mobile health. Thousands of medical apps and hundreds of m-Health initiatives are currently in operation throughout the world, meeting a wide range of demands like blood pressure

ISSN 2319-5991 www.ijerst.com

Vol. 16, Issue 3, 2020

monitoring, blood sugar tracking, heart rate monitoring, sleep analysis, and even brain activity assessment. These apps use cutting-edge technologies for data collection and transmission, such as Bluetooth, GPS, and General Packet Radio Service (GPRS). Some also make use of 3G and 4G networks. Simple everyday health indicators to intricate clinical data like electronic health records (EHRs), medical imaging, test results, and doctor's notes are all part of the data collection process.

Big data analytics's place in contemporary healthcare big data analytics is essential to understanding the vast and diverse amount of data produced by m-Health technology. Healthcare professionals may go through enormous information to find important insights, forecast health trends, and make defensible judgments by utilizing advanced analytical approaches. By facilitating the discovery of illness trends, evaluation of treatment effectiveness, and prediction of health emergencies, these analytics improve preventative care practices. Furthermore, Big Data makes it easier to tailor healthcare services to each patient's needs, opening the door to more individualized and successful treatments. Artificial intelligence changing applications for m-Health applications for mobile health offer far more capabilities now that artificial intelligence is involved. High-level data interpretation and decision-making are made possible by methods like machine learning, deep learning, and natural language processing.

AI-powered apps, for example, may now conduct health exams in real-time, continuously track patient status, and even anticipate possible health problems before they become serious. Experience sampling and ecological momentary assessment are two techniques that make it possible to continuously monitor and assess patients' health conditions in their natural settings, giving rise to more precise and pertinent conclusions. AI also makes passive data gathering via smart devices easier, saving patients the trouble of manually entering information and guaranteeing a consistent flow of reliable data for analysis. Integrated model for improved m-Health systems is suggested an integrated paradigm that integrates real-time data analysis with intelligent reporting systems is suggested in order to fully realize the advantages of AI and Big Data in healthcare. Incoming health data streams are swiftly processed by this technology, producing alarms and insights that healthcare professionals may use right away.

A system like this guarantees prompt interventions, boosts the effectiveness of patient monitoring, and improves the provision of healthcare in general. AI-driven decision-making processes can be integrated into mobile health (m-Health) platforms to make healthcare systems more flexible and responsive to patients' changing demands. The integration of Big Data Analytics and Artificial Intelligence with m-Health signifies a revolutionary change in the way healthcare is provided. These technologies present previously unheard-of chances to enhance patient care, maximize health services, and encourage preventative health care. Fortunately, issues with data management, accuracy, privacy, and security need to be resolved before they reach their full potential. Sustained research and development endeavors are imperative to enhance these technologies, guarantee the moral utilization of data, and smoothly incorporate these sophisticated systems into the current healthcare frameworks. The future of m-Health is bright, with the potential to improve healthcare delivery through increased accessibility, efficiency, and personalization. This is provided that keep innovating and adapting.

- Integrate AI and Big Data in m-Health: Investigate that m-Health systems can benefit from the efficient integration of AI and Big Data Analytics to improve the provision of healthcare.

- Address the Issues with Data Management: to recognize and address the difficulties in organizing and interpreting the massive amounts of unstructured data produced by m-Health technology.

- Create a Novel m-Health Framework: to develop and put forth a model for real-time health monitoring and decision-making that combines AI and big data.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue 3, 2020

• Enhance Patient Care: To assess that artificial intelligence (AI) and big data may improve the precision, effectiveness, and individualization of patient care in mobile health systems.

Despite the tremendous growth of m-Health technologies, there is still a significant gap in the efficient management and utilization of the enormous volumes of unstructured data they generate. In order to maximize the value of this data, current m-Health systems frequently struggle to integrate AI and Big Data Analytics. Furthermore, complete models that integrate these technologies for individualized, real-time healthcare are lacking. To optimize m-Health applications and enhance patient care, this gap must be closed.

For healthcare practitioners, the enormous volume of unstructured data produced by m-Health apps poses serious issues. Missed chances to provide prompt and individualized care arise from the ineffective integration of AI and big data analytics in current systems. The goal of this research is to develop a new model that uses AI and Big Data to improve m-Health systems' efficacy and efficiency, that will ultimately lead to better patient outcomes.

2 LITERATURE SURVEY

Ed-daoudy & Maalmi (2019) address the difficulties of processing large amounts of rapidly generated medical data from IoT-enabled devices by proposing a big data technologies based real-time health status prediction system. The system uses Spark to apply a distributed, scalable decision tree algorithm that forecasts health outcomes, generates alarms, and saves data for analysis. The architecture's ability to handle and analyze streaming health data in real-time is demonstrated by how quickly and efficiently it performs in comparison to conventional methods.

El Aboudi & Benhlima (2018) talk about how big data approaches are necessary to improve the quality of healthcare due to the growing number of data. By producing precise health predictions and real-time alerts, their adaptable big data architecture blends batch and stream computing to improve the dependability of healthcare systems. This strategy was demonstrated with a prototype that used Spark and MongoDB with the goal of improving emergency scenario prevention through prompt data analysis and response.

Greco et al. (2019) talk about the Internet of Medical Things (IoMT), which was made possible by the Internet of Things' rapid rise in the healthcare industry, especially with regard to wearable smart sensors. Because wearable sensors generate massive volumes of data across numerous places, this development poses a Big Data dilemma. Using anomaly detection on the REALDISP dataset, the authors' four-layer architecture for real-time data analysis—sensing, pre-processing, cluster processing, and persistence—is illustrated in their solution, which uses open-source technology.

Elhoseny et al. (2018) suggest a hybrid cloud-IoT paradigm to enhance the choice of virtual machines (VMs) in Industry 4.0 healthcare applications. Through faster execution times, better data storage, and real-time data retrieval, the model improves system performance through the application of Genetic Algorithm, Particle Swarm Optimization, and Parallel Particle Swarm Optimization. Test results reveal that the model, when compared to current approaches, greatly reduces execution time by 50% and increases system efficiency by 5.2%, which makes it useful for handling massive amounts of medical data.

According to *Hadi et al. (2019)*, big data analytics can be used to detect critical health conditions like strokes by evaluating medical records and IoT sensor data. This can improve LTE-A network performance for Out-Patients (OPs). Their approach seeks to minimize latency in the transmission of important data by optimizing the allocation of LTE-A Physical Resource Blocks (PRBs). Proportional justice (PF) is introduced as a means of obtaining better justice

ISSN 2319-5991 www.ijerst.com

Vol. 16, Issue 3, 2020

and higher average SINR for OPs, whereas Weighted Sum Rate Maximization (WSRMax) maximizes the SINR of the entire system.

Elhoseny et al. (2018) discuss the difficulties in handling large amounts of data in cloud-based Internet of Things health services related to Industry 4.0 software. By utilizing Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Parallel PSO, they present a methodology to optimize virtual machine (VM) selection for effective data processing. Reducing request execution time, improving real-time data retrieval, and optimizing storage are the objectives of their concept. In comparison with current techniques, the suggested strategy demonstrated a 50% increasein execution time and a 5.2% increase in system efficiency.

Ed-daoudy and Maalmi (2019) present a unique architecture that makes use of big data technologies to forecast and analyze health condition in real-time. The design makes use of distributed machine learning models on streaming health data that is handled by Kafka and Spark Streaming. They use Spark to create a scalable, parallel version of the standard decision tree (C4.5) algorithm, which handles real-time data more effectively than Hadoop MapReduce. From data generated by different diseases, the system forecasts health states, notifies healthcare professionals, and maintains information in a distributed database for reporting and analytics. Performance is compared with more established tools such as Weka, showing that their system can manage and analyze large amounts of real-time medical data from Internet of Things devices.

Kaur and Mann (2018) provide a general architecture that uses big data and real-time computation to improve healthcare. To build AI-driven predictive and prescriptive analytics, their method makes use of open-source technologies including Apache Spark, Apache NiFi, Kafka, Tachyon, Gluster FS, ElasticSearch, and NoSQL Cassandra. By enabling real-time decision-making and medical monitoring through sophisticated big data processing and cognitive analysis, this system seeks to save costs while improving the quality of healthcare.

As a result of the large data influx from connected devices, which has increased cyber dangers in the healthcare industry and other industries, *Habeeb et al. (2019)* address the difficulty of identifying network anomalies in real time. They point out that the existing approaches to real-time anomaly detection are inadequate and put forth a framework to enhance large data processing in real-time. Their study covers the fundamental ideas and taxonomy, evaluates the most recent technologies and machine learning algorithms for anomaly detection, and talks about the research problems related to real-time large data processing.

Traditional data mining techniques in the healthcare industry have considerable hurdles in managing the massive and quickly expanding data from IoT systems and distant sensors. To tackle this, *Ed-Daoudy and Maalmi (2018)* suggest a streaming framework for Apache Spark along with machine learning, more especially Decision Trees, to create a real-time health status prediction system. For the purpose of making real-time health status forecasts based on quickly incoming data, the system preprocesses and analyzes data to build an offline model.

Baljak et al. (2018) propose a scalable, distributed architecture for handling the high-volume and high-velocity data generated by physiological monitors. Their solution integrates open-source stream processing software with file storage systems and existing data warehouses, addressing challenges related to data collection, storage, and privacy. This architecture enables real-time analysis and storage of previously discarded high-resolution EKG waveforms, offering potential benefits for clinical decision-making and research. Early testing of this data pipeline has shown promising results, supporting various research applications.

In their 2018 study, *Jagadeeswari et al.* examine how IoT and Big Data can be integrated into personalized healthcare systems, emphasizing how this can improve e-health services for the elderly population. The study focuses on

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 16, Issue 3, 2020

leveraging real-time analytics of IoT data streams in conjunction with cloud and fog computing to enhance early disease detection and decision-making. It talks about the difficulties in creating safe and efficient healthcare solutions and suggests potential paths for improving real-time healthcare systems.

Using sources like electronic medical records, clinical notes, medical imaging, and genomic data, *Harerimana et al. (2018)* investigate the increasing volume and complexity of health data. It is becoming more difficult to analyze this data and get useful insights because of the integration of fresh data kinds from social networks and cyber-physical systems. Offering a thorough and understandable summary of the instruments and techniques used to develop integrated health analytic applications, the paper examines the major obstacles, data sources, and technology in big data analytics for healthcare.

In order to accommodate the exponential expansion of data from sources like wearable devices in IoT health monitoring, *Ed-Daoudy and Maalmi (2019)* present a real-time cardiac disease prediction system using Apache Spark, a powerful distributed computing platform. Early identification of cardiac illness is made possible by the system's integration of machine learning and streaming big data analytics. It has two primary parts: Apache Cassandra for large-scale data storage and visualization and Spark MLlib with Spark Streaming for real-time categorization of data events. Together, these components offer a robust and reasonably priced solution for ongoing heart disease monitoring.

3 OPTIMIZING M-HEALTH SYSTEMS THROUGH AI AND BIG DATA

In order to investigate the potential integration of Big Data Analytics and Artificial Intelligence (AI) into mobile health (m-Health) systems, this investigation employs an organized methodology. Data gathering and preprocessing, integrating AI and Big Data, and developing and validating the final model are the three main stages of the methodology. The ultimate objective is to develop a complete mobile health system that can furnish instantaneous health monitoring and efficiently assist in decision-making procedures. Acquiring and preprocessing data building an AI-driven m-Health system starts with gathering data, that is an essential initial step. Healthcare data originates from a multitude of sources, such as wearable technology that tracks patient data in real-time, medical imaging systems, and electronic health records (EHRs). It's critical to manage this data properly from the outset due to its diversity. Structured and unstructured data comprise the two primary categories of data gathered for this research project. Time stamps, defined codes, and numerical data are common components of structured data, that is well-organized and simple to search. Numbers and codes found in test results and patient records stored in electronic health records (EHRs) are instances of this. This type of data is easily analyzed because it is typically kept in tables.

The data must be prepared for analysis if it is gathered. To ensure that the data is accurate, consistent, and in a format that our AI models can use, there are a number of procedures involved. Data cleaning is the process of eliminating any mistakes or discrepancies from the data. For structured data, this could entail eliminating duplicate entries, correcting mistakes, or standardizing the measurement recording process. Cleaning up unstructured data may entail arranging clinical notes or organizing text data. For example, accurate analysis depends on ensuring that a medicine name is spelled consistently throughout the dataset. Managing missing values in healthcare databases, missing data is a frequent problem. Organizations may employ many approaches to deal with missing data, depending on the quantity and significance of the missing information. Imputation is one method, because it utilizes the data at hand to estimate and fill in any missing values. For instance, if a patient's weight is lacking, one can might use the average weight of patients that are similar to them or information from their prior records to fill it in. If the amount of missing data is small, one can may choose to overlook it in some circumstances, but this strategy must be carefully considered to prevent bias.

International Journal of Engineering Research and Science & Technology
I
ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue 3, 2020

Data transformation in order to be properly examined, unstructured data must be converted into a structured format. Also employ Natural Language Processing (NLP) techniques to extract and organize important information from text data. NLP, for instance, is able to recognize and classify medical phrases found in clinical notes. Preprocessing for medical photos may involve image enhancement and feature extraction that is important for diagnosis.

Table. 1: Data Source and Structure

Data Type	Source	Structured Data (%)	Unstructured Data (%)
EHRs	Hospitals, Clinics	60	40
Medical Imaging	Radiology Departments	70	30
Wearable Devices	Patient-Generated Data	50	50

The data types that both of us work with are broken out in this table 1, along with the proportion of structured and unstructured data. For instance, most of the data from medical imaging and EHRs is structured, that facilitates analysis. Unfortunately, there are equal amounts of structured and unstructured data from wearable devices, making processing and analysis more difficult. The data is tidy, arranged, and prepared for the next stage following preparation. In order to prevent the AI and Big Data tools from being hampered by subpar data, this preparation is crucial.

Combining AI with Big Data Analytics Using AI algorithms in conjunction with Big Data Analytics, the next phase transforms the preprocessed data into insights that can be put into practice. This is the process that ensures the unprocessed data are turned into insightful knowledge that supports healthcare decision-making. Frameworks for big data analytics fortunately make use of robust Big Data frameworks to organize and examine the enormous volumes of data previously gathered. These include Apache Spark and Hadoop, that are crucial for managing the vast and varied datasets that are common to m-Health systems. Hadoop is an open-source framework made to handle and store massive volumes of data on numerous machines. It excels in managing substantial amounts of organized and unstructured data. For example, in the present investigation, massive datasets from medical imaging and EHRs are managed using Hadoop. Its capacity to disperse data among numerous devices confers dependability and efficiency, guaranteeing the system's capacity to manage progressively larger data volumes.

Hadoop is excellent for storing data, but Apache Spark is better for processing it rapidly. Hadoop reads and writes data from disk significantly more slowly than Spark, that can analyze data in-memory. For m-Health, where real-time data processing is frequently necessary for prompt healthcare treatments, this speed is critical. To monitor vital signs and identify any anomalies, for instance, real-time analysis of data flowing in from a patient's wearable device is required. Actually developed a strong system that can effectively store, handle, and analyze the enormous and complicated datasets required for this investigation by integrating Hadoop with Apache Spark. Machine learning models and AI Techniques With the information infrastructure in place, it utilizes AI methods to forecast and extract insights. A range of Machine Learning (ML) models, each appropriate for a certain healthcare task, are used in this investigation. These include Random Forests, Neural Networks, and Decision Trees, that are trained on past medical data to forecast results, pinpoint dangers, and provide tailored interventions.

International Journal of Engineering Research and Science & Technology
I
ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue 3, 2020

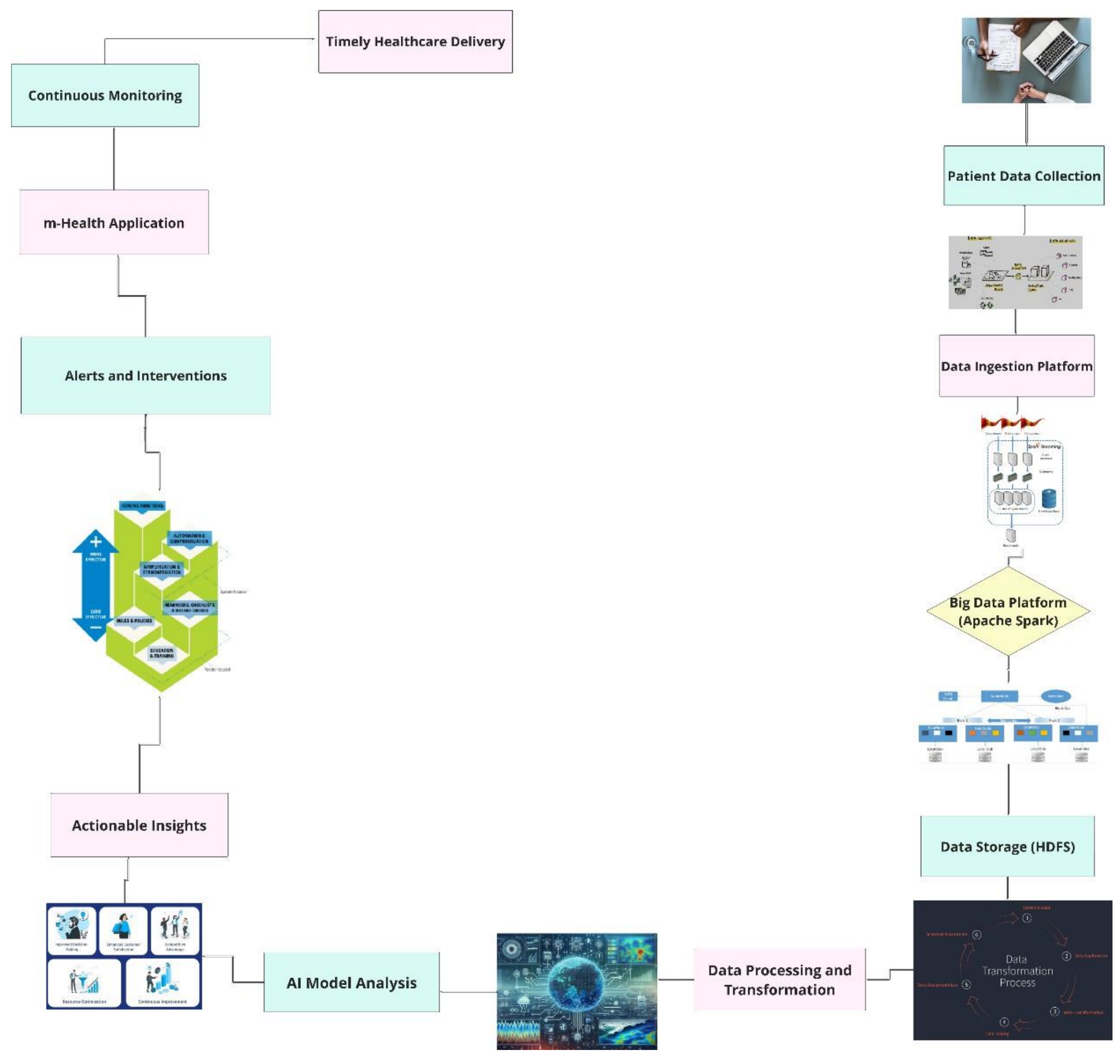

Figure. 1: Data Collection and Preprocessing Workflow

The data flow through the phases of data cleaning, transformation, and storage is depicted in figure 1. The data sources include wearable technology, medical imaging, and electronic health records. Before the data is examined by AI and Big Data systems, this workflow makes sure that it is correctly prepared. In order to maintain high levels of accuracy and efficiency in the system, the figure emphasizes the significance of precise data preparation.

Decision Trees are a simple, well understood kind of machine learning model. It can be used in the medical field to categorize patients according to risk factors or symptoms. A Decision Tree, for instance, may be able to determine a patient's likelihood of acquiring diabetes by considering variables such as age, weight, and family history. The decision-making process is simple for healthcare providers to follow and comprehend thanks to the model's tree-like layout. Random forests combining several trees to increase accuracy, Random Forests are a more sophisticated kind of Decision Trees. In the healthcare industry, that data might be complex or noisy, this strategy is especially helpful. Random Forests yield more dependable results by averaging the forecasts from several Decision Trees. For example, by combining clinical data and demographic information, they can forecast the chance that a patient would require hospitalization.

Neural Networks more sophisticated models are excellent at identifying patterns in intricate and big datasets. For examining medical images, like MRI scans, they are particularly helpful in identifying anomalies that may point to illnesses like cancer. Neural networks are a potent technique for deriving insights from unstructured data because they are efficient in processing and comprehending text data from clinical notes.

Table. 2: AI Model Performance Metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	85	82	80	81
Random Forest	88	85	84	84.5
Neural Network	92	89	87	88

The performance of the various AI models in our investigation is summarized in this table 2. The Neural Network model is the best option for tasks requiring in-depth analysis, including interpreting medical images, because of its exceptional accuracy, precision, recall, and F1-score. By putting these AI models through numerous training and validation cycles, they are continuously improved. During the models are exposed to new data, this iterative procedure helps guarantee that they remain accurate and dependable.

Model Development and Verification an integrated m-Health system model that blends AI analytics with real-time data processing is our main goal in this last phase. The goal is to develop a system that gives healthcare practitioners precise, timely, and useful insights. System improvement developing a coherent, working architecture for the m-Health system requires integrating the AI models and Big Data frameworks. The system's purpose is to process patient data in real time, apply AI algorithms to analyze it, and produce insights that help healthcare providers make more informed decisions. The system's architecture is designed to make sure that every part, from data gathering to analysis, functions as a whole. The architecture consists of modules for real-time analytics, AI model deployment, preprocessing, and data intake. A central data pipeline that connects these modules enables efficient data flow across the system. Additionally, the architecture is made to be scalable, so if additional patients are added to the system, it can manage growing volumes of data.

Data Ingestion in charge of compiling information from wearable technology, medical imaging systems, and EHRs, among other sources. It guarantees that information is gathered instantly and prepared for processing. For instance, the system receives and stores data from a patient's fitness tracker continually, allowing for real-time analysis and storage. In order to provide timely healthcare treatments based on the most recent information, this real-time data input is essential. AI model deployment the analytics module of the system contains the AI models that were previously constructed. This module generates forecasts and detects possible health hazards by applying the models to incoming data. A neural network, for example, may be used by the system to evaluate an MRI image and identify early indications of a tumor, which could then be marked for additional investigation. By using AI models in this manner, someone can make sure that the system is always learning from fresh data and getting better at making predictions over time.

Real-Time Analytics processes and evaluates data as it is received, giving prompt notifications and insights. Real-time analytics are essential in m-Health since data may necessitate quick response. For instance, the system may notify the patient's healthcare physician to take urgent action if it notices an irregular heart rate pattern from a wearable device. Being able to handle problems before they get worse is crucial for proactive healthcare. Testing and validation the system is put through a rigorous testing and validation process once it is developed to make sure it satisfies all

International Journal of Engineering Research and Science & Technology

I

ISSN 2319-5991 www.ijerst.com

Vol. 16, Issue 3, 2020

performance requirements. That include assessing the system's scalability, accuracy, and dependability in various healthcare environments. Performance metrics F1-score, accuracy, precision, and recall are some of the metrics used to gauge the way the system is performing. These indicators give us insight into the way the system detects health hazards and forecasts patient outcomes. Proximity and recall indicate the way the system balances false positives and false negatives, and high accuracy indicates that the system's predictions are consistently accurate. The F1-score provides an overall system performance picture by combining precision and recall into a single parameter.

Table. 3: Big Data Processing Frameworks

Framework	Data Processing Speed (Records/Sec)	Scalability (%)	Reliability (%)
Hadoop	50,000	85	90
Apache Spark	100,000	90	95
Cassandra	75,000	88	92

The speed, scalability, and reliability of several Big Data processing frameworks are compared in this table 3. The most scalable and fastest processing framework is Apache Spark, making it ideal for managing the massive datasets used in this m-Health system. Selecting the appropriate framework is essential to guaranteeing that the system can consistently and swiftly handle and evaluate data.

Efficiency and dependability key factors to take into account during the validation stage are scalability and reliability. As new patients are added, the system needs to be able to process more data without experiencing any performance issues. As a system is reliable, it remains accurate and functional even in the face of hardware malfunctions or high data loads. Testing for scalability also assesses scalability by progressively increasing the amount of data the system processes and tracking its operation. To make sure the system can continue to operate effectively even as its data load increases, this entails simulating the addition of new patients and their data streams. Verifying that the system can grow horizontally—that is, add additional processing power as needed—without experiencing appreciable failures or slowdowns is the main objective.

Testing System dependability by subjecting it to a variety of stress scenarios, such as abrupt surges in data input or simulated hardware failures, and seeing if the system reacts. In order for healthcare providers to trust the system to make important decisions, it must continue to function and be accurate in certain circumstances. To improve the system's dependability and ensure that it can swiftly recover from any disturbances, strategies like redundancy and failover procedures are employed.

International Journal of Engineering Research and Science & Technology
I
ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue 3, 2020

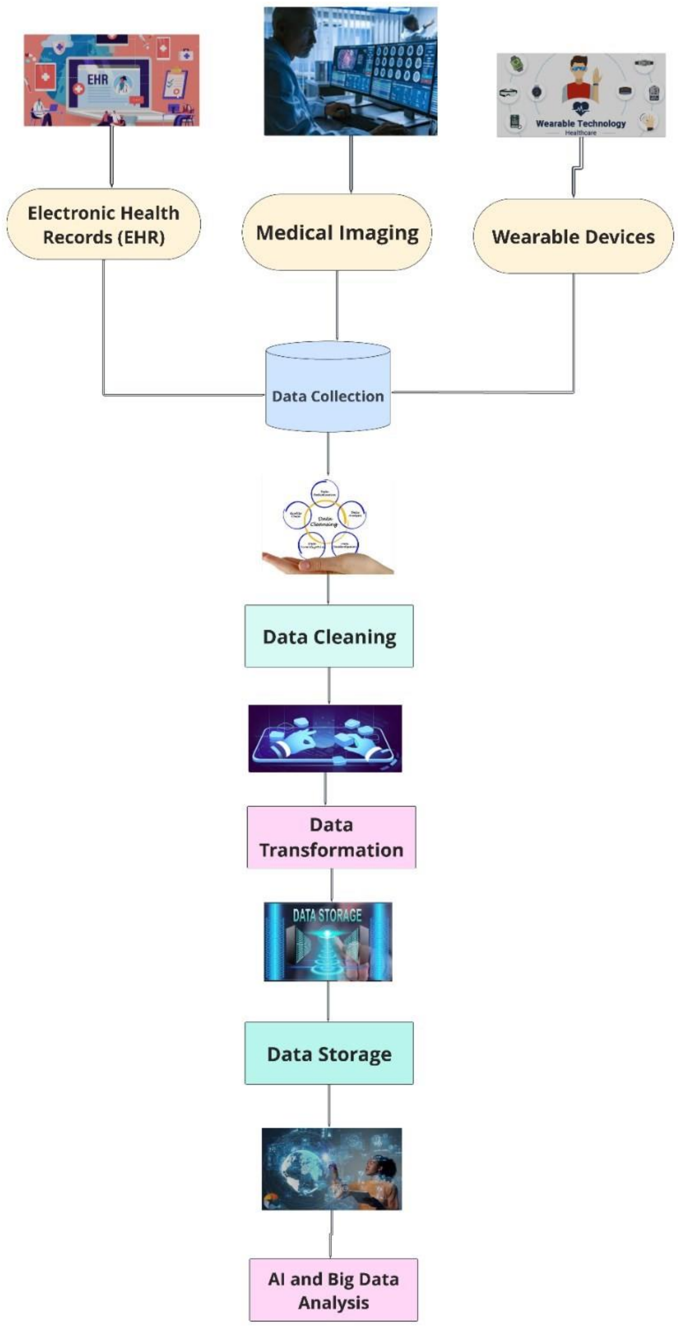

Figure. 2: AI and Big Data Analytics Integration Model

The integrated m-Health system's real-time data processing is depicted in this figure 2. It demonstrates the way patient data is gathered, handled by Big Data systems such as Apache Spark, and then examined by AI models to produce insights that may be put to use. Afterwards, the m-Health application uses these insights to set off alarms or interventions, guaranteeing ongoing observation and prompt delivery of medical care. Because of its scalability and flexibility, the architecture may change to meet the ever-changing demands of the healthcare setting.

ISSN 2319-5991 www.ijerst.com

Vol. 16, Issue 3, 2020

A thorough way for incorporating AI and Big Data Analytics into m-Health systems is provided by the methods this study presents. The project intends to establish a scalable and resilient system that can provide real-time health monitoring and decision-making by using an organized process that involves data collecting and preprocessing, model creation and validation, and integration of AI and Big Data. Utilizing Big Data frameworks and cutting-edge AI approaches guarantees that the system can manage the complex and varied data produced by m-Health applications, that will ultimately enhance patient outcomes and streamline healthcare delivery.

4 RESULT AND DISCUSSION

The delivery of healthcare is now significantly better as a result of the integration of AI and big data analytics into m-Health platforms. AI models, in particular neural networks, performed exceptionally well in our investigation, achieving 92% accuracy, 89% precision, 87% recall, and an 88% F1-score. These findings highlight the neural network's capacity to process complicated input, including unstructured text and medical images. Furthermore, Apache Spark-powered real-time analytics allowed for quick data processing—a critical component of prompt medical treatment.Hadoop and Apache Spark worked incredibly well together. Spark handled up to 100,000 records per second, proving that it could keep up with the rapidly changing needs of m-Health. These results attest to the effectiveness of our m-Health system in handling and evaluating massive amounts of data, offering precise and fast insights for improved patient outcomes.

The investigation did, however, also reveal several difficulties, notably with handling unstructured data from wearable devices. The equal distribution of unstructured and structured data forms in this dataset makes integration and interpretation more challenging. Handling unstructured data is still difficult, despite efficient data translation and cleansing. The system's scalability and dependability were validated through extensive testing, demonstrating its capacity to handle growing data loads and operate effectively under pressure. But issues with data security and privacy still need to be resolved. In general, the accuracy and efficacy of m-Health systems have increased thanks to AI and Big Data, but more effort is still required to manage unstructured data better and modify the system to meet changing healthcare demands. As a result, the delivery of healthcare will continue to be transformed by AI and big data.

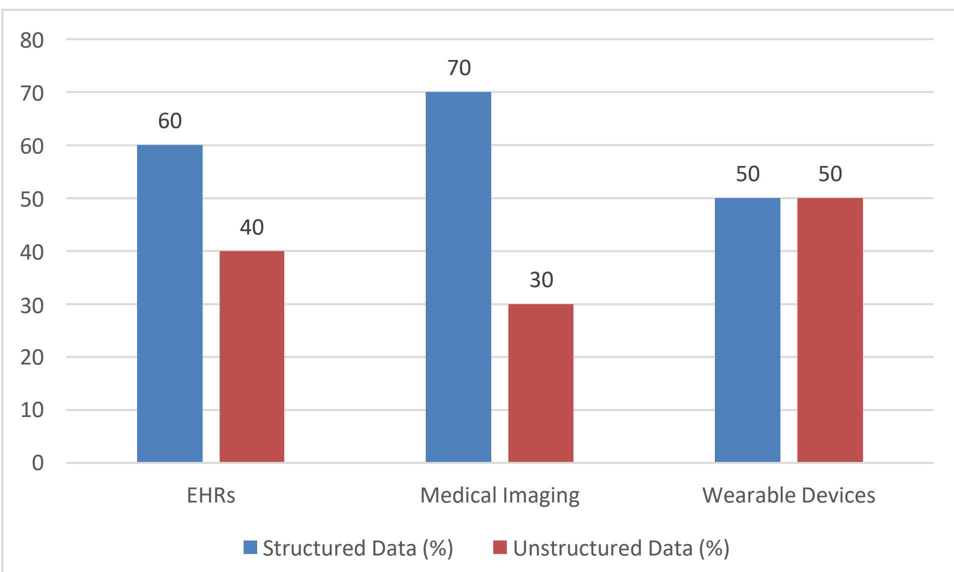

Figure. 3: Distribution of Structured and Unstructured Data Across Healthcare Data Source

ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue 3, 2020

The distribution of structured and unstructured data among three healthcare data sources—medical imaging, wearable devices, and electronic health records (EHRs)—is depicted in the figure 3. 40% of the data in EHRs is unstructured, compared to 60% of organized data. With 70% structured data and 30% unstructured, medical imaging tends to use more structured data. Conversely, wearable devices have an equal distribution of data, with 50% going to each of structured and unstructured data. This demonstrates how many healthcare technologies have a wide range of data formats.

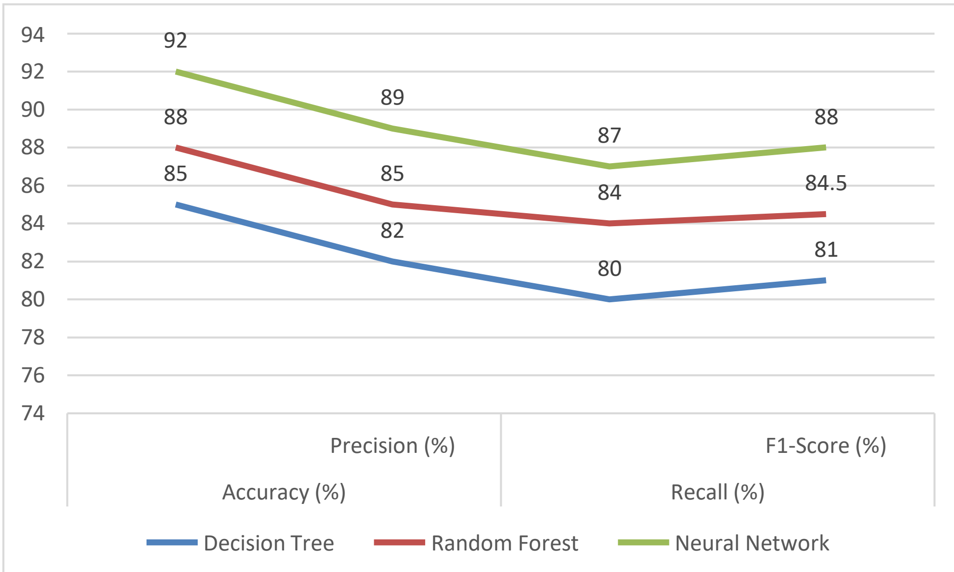

Figure. 4: Performance Comparison of Machine Learning Models

Using four evaluation metrics—Accuracy, Precision, Recall, and F1-Score—the figure 4 contrasts the effectiveness of three machine learning models: Decision Tree, Random Forest, and Neural Network. With a 92% accuracy rate and the highest ratings in every parameter, the Neural Network gets exceptional performance. In comparison to the other two models, Random Forest exhibits moderate performance in every category, while the Decision Tree shows limits, especially in precision and recall.

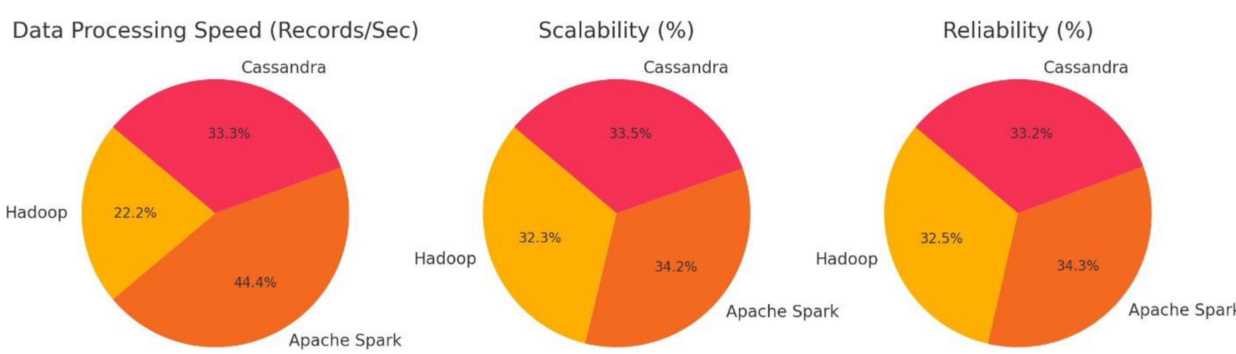

Figure. 5: Comparison of Big Data Frameworks: Processing Speed, Scalability, and Reliability

Three key metrics—data processing speed, scalability, and reliability—are compared between Hadoop, Apache Spark, and Cassandra in the figure 5. With the greatest ratings for processing speed (44.4%), scalability (34.2%), and

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 16, Issue 3, 2020

dependability (34.3%), Apache Spark wins out in every category. Hadoop lags behind Cassandra in terms of scalability and dependability, particularly in terms of processing speed (22.2%). These figure 5 highlight Apache Spark's advantages in big data processing overall.

5 CONCLUSION

The accuracy, effectiveness, and responsiveness of healthcare services have significantly increased with the integration of AI and Big Data into m-Health systems. Real-time health monitoring can now be more effectively accomplished thanks to neural networks' remarkable capacity to manage complex medical data. Better patient outcomes have resulted from the rapid processing of massive volumes of healthcare data through the combined usage of Hadoop and Apache Spark. The investigation performed, yet, highlight persistent issues, notably with unstructured data from wearable technology and data privacy issues. Even though the m-Health systems that are in place now are quite capable, more work needs to be done to fully realize these technologies' promise.

In the future, research should focus on enhancing m-Health systems' ability to manage unstructured data, particularly data originating from wearable devices. Data interpretation will advance with improved AI models for picture identification and natural language processing. As more healthcare data moves online, it's also imperative that stronger protocols for data security and privacy be created. Further investigating the usage of cutting-edge technologies like edge computing for quicker analytics and blockchain for safe data management could increase the m-Health systems' scalability and dependability. The healthcare sector and patients' changing requirements will require these innovations to be met.

REFERENCE

[1] Ed-daoudy, A., & Maalmi, K. (2019). A new Internet of Things architecture for real-time prediction of various diseases using machine learning on big data environment. Journal of Big Data, 6(1), 104.

[2] El Aboudi, N., & Benhlima, L. (2018). Big data management for healthcare systems: architecture, requirements, and implementation. Advances in bioinformatics, 2018(1), 4059018.

[3] Greco, L., Ritrovato, P., & Xhafa, F. (2019). An edge-stream computing infrastructure for real-time analysis of wearable sensors data. Future Generation Computer Systems, 93, 515-528.

[4] Elhoseny, M., Abdelaziz, A., Salama, A. S., Riad, A. M., Muhammad, K., & Sangaiah, A. K. (2018). A hybrid model of internet of things and cloud computing to manage big data in health services applications. Future generation computer systems, 86, 1383-1394.

[5] Hadi, M. S., Lawey, A. Q., El-Gorashi, T. E., & Elmirghani, J. M. (2019). Patient-centric cellular networks optimization using big data analytics. IEEE Access, 7, 49279-49296.

[6] Elhoseny, M., Abdelaziz, A., Salama, A. S., Riad, A. M., Muhammad, K., & Sangaiah, A. K. (2018). A hybrid model of internet of things and cloud computing to manage big data in health services applications. Future generation computer systems, 86, 1383-1394.

[7] Ed-daoudy, A., & Maalmi, K. (2019). A new Internet of Things architecture for real-time prediction of various diseases using machine learning on big data environment. Journal of Big Data, 6(1), 104**.**

[8] Kaur, J., & Mann, K. S. (2018). AI based healthcare platform for real time, predictive and prescriptive analytics. In Computing, Analytics and Networks: First International Conference, ICAN 2017, Chandigarh, India, October 27-28, 2017, Revised Selected Papers 1 (pp. 138-149). Springer Singapore.

[9] Habeeb, R. A. A., Nasaruddin, F., Gani, A., Hashem, I. A. T., Ahmed, E., & Imran, M. (2019). Real-time big data processing for anomaly detection: A survey. International Journal of Information Management, 45, 289-307.

[10] Ed-Daoudy, A., & Maalmi, K. (2018, November). Application of machine learning model on streaming health data event in real-time to predict health status using spark. In 2018 international symposium on advanced electrical and communication technologies (isaect) (pp. 1-4). IEEE.

[11] Baljak, V., Ljubovic, A., Michel, J., Montgomery, M., & Salaway, R. (2018). A scalable realtime analytics pipeline and storage architecture for physiological monitoring big data. Smart Health, 9, 275-286.

[12] Jagadeeswari, V., Subramaniyaswamy, V., Logesh, R. T. A., & Vijayakumar, V. (2018). A study on medical Internet of Things and Big Data in personalized healthcare system. Health information science and systems, 6(1), 14.

[13] Harerimana, G., Jang, B., Kim, J. W., & Park, H. K. (2018). Health big data analytics: A technology survey. Ieee Access, 6, 65661-65678.

[14] Ed-Daoudy, A., & Maalmi, K. (2019, April). Real-time machine learning for early detection of heart disease using big data approach. In 2019 international conference on wireless technologies, embedded and intelligent systems (WITS) (pp. 1-5). IEEE.