EHR VAULT

CONFIDENTIAL
HEALTH REPOSITORY

EHR VAULT

A Secure and Anonymized Framework for Electronic Health Records Data Warehousing

Guided by
Dr. Kathiroli R

Vijai Suria M,
2021503568
Dept. of Computer Technology, MIT

Thanes M
2021503712
Dept. of Computer Technology, MIT

Sri Ram Kumar G
2021503322
Dept. of Computer Technology, MIT

Motivation

- **Addressing security vulnerabilities:** Traditional centralized approaches to EHR management are prone to security breaches, raising concerns about patient privacy and data integrity.
- **Enhancing scalability and interoperability:** Existing EHR systems may struggle to scale with growing data volumes and lack interoperability between different healthcare providers.
- **Improving decision-making and patient care**: By integrating IPFS for decentralized storage and data warehousing for centralized analysis, our framework enables comprehensive analytics and reporting, leading to better decision-making and patient care outcomes.

LITERATURE SURVEY

Paper Name	Journal / Year	Methodology / Findings	Limitations
[1] Federated Learning-Based Secure Electronic Health Record Sharing Scheme in Medical Informatics.	IEEE Journal of Biomedical and Health Informatics (2023).	Involves implementing a Federated Learning-based, CNN model and private InterPlanetary File System (IPFS) storage for medical data security.	Potential challenges related to scalability and adoption of the proposed Federated Learning-based Electronic Health Record sharing scheme across diverse healthcare settings.
[2] Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis.	IEEE J Biomed Health Inform (2018).	Research on applying deep learning to clinical tasks based on EHR data.	Challenges related to model interpretability, data heterogeneity, may hinder the widespread adoption and generalizability of deep learning techniques in clinical informatics applications based on EHR data.

LITERATURE SURVEY

Paper Name	Journal / Year	Methodology / Findings	Limitations
[3] Flexible Data Warehouse: Towards Building an Integrated Electronic Health Record Architecture.	International Conference on Smart Electronics and Communication (ICOSEC) (2020).	Integration of flexible data warehouse technology into electronic health record systems.	The integration of data warehousing technologies in electronic health record systems may face challenges in achieving higher performance and effectiveness.
[4] A Knowledge Distillation Ensemble Framework for Predicting Short- and Long-Term Hospitalization Outcomes From Electronic Health Records Data.	IEEE Journal of Biomedical and Health Informatics (2022).	Utilizing an unsupervised LSTM Autoencoder, combined with a gradient boosting mode, EHR data.	Lack of generalizability beyond ICU and potential biases in the dataset used for training and evaluation.

LITERATURE SURVEY

Paper Name	Journal / Year	Methodology / Findings	Limitations
[5] Efficient Mining Template of Predictive Temporal Clinical Event Patterns From Patient Electronic Medical Record.	IEEE Journal of Biomedical and Health Informatics (2019).	Mining patterns in EMR data, followed by evaluation of features in predictive modeling for early detection of congestive heart failure.	Potential challenges in generalizing findings to other disease conditions and datasets, and limitations in capturing complex temporal dependencies beyond point events in EMR data.
[6] A Cloud-based Approach for Interoperable Electronic Health Records (EHRs).	IEEE Journal of Biomedical and Health Informatics (2018).	Semantic interoperability in the proposed cloud-based EHR system (CHISTAR).	CHISTAR face challenges in achieving widespread adoption due to data security and privacy in cloud-based EHR systems.

LITERATURE SURVEY

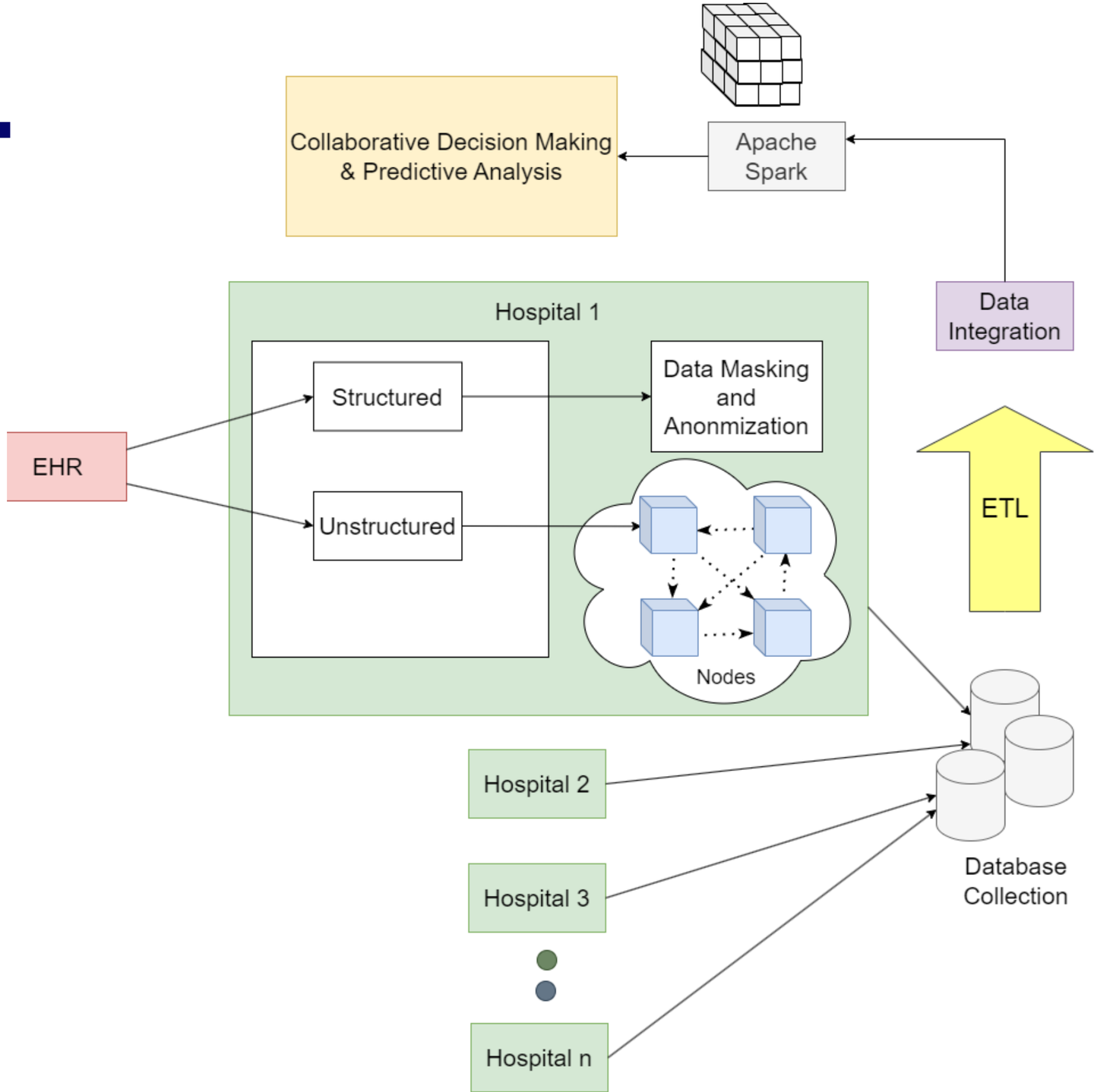
Paper Name	Journal / Year	Methodology / Findings	Limitations
[7] Hi-BEHRT: Hierarchical Transformer-Based Model for Accurate Prediction of Clinical Events Using Multimodal Longitudinal Electronic Health Records.	IEEE Journal of Biomedical and Health Informatics (2023).	Hierarchical Transformer-based model.	Potential challenges related to interpretability of the hierarchical Transformer model and generalizability to diverse healthcare settings and patient populations.
[8] A Shared Decision-Making System for Diabetes Medication Choice Utilizing Electronic Health Record Data.	IEEE Journal of Biomedical and Health Informatics (2017).	SDM system framework for T2DM patients, multilabel classification model using class-imbalanced EHR data.	Potential challenges related to model generalizability to diverse patient populations and settings, and limitations in capturing complex patient preferences and individual health contexts.

LITERATURE SURVEY

Paper Name	Journal / Year	Methodology / Findings	Limitations
[9] A Regularized Deep Learning Approach for Clinical Risk Prediction of Acute Coronary Syndrome Using Electronic Health Records.	IEEE Transactions on Biomedical Engineering (2018).	Stacked denoising auto-encoder (SDAE) enhance feature representations for improved risk prediction.	Potential challenges related to generalizability of the model to diverse patient populations and settings, and limitations in capturing all relevant risk factors for ACS prediction.
[10]HCET: Hierarchical Clinical Embedding With Topic Modeling on Electronic Health Records for Predicting Future Depression.	IEEE Journal of Biomedical and Health Informatics (2021).	HCET utilize heterogeneous EHR data for predicting depression, with attention mechanisms.	Potential challenges related to generalizability of the model beyond depression prediction and scalability to larger and more diverse datasets.

SYSTEM MODEL

- Cutting-edge integration: The "EHR Vault" framework merges IPFS and data warehousing for EHR management.
- Decentralized storage: Branches host IPFS nodes for secure data storage and collaboration.
- Centralized analytics: Data from IPFS nodes undergo processing in a centralized warehouse for scalable analytics.



Objectives

- Implementing a secure and decentralized framework for EHR data management.
- Enhancing scalability and interoperability of EHR systems.
- Enabling comprehensive analytics and reporting for informed decision-making.
- Empowering healthcare organizations with data-driven insights for improved patient care.

IMPLEMENTATION

- Deploying IPFS nodes at each branch for decentralized storage of EHR data.
- Integrating traditional RDBMS (e.g., MySQL) at local branches for structured data storage

- Developing synchronization mechanisms to replicate EHR data from local databases to the centralized data warehouse.

- Utilizing ETL processes or real-time data streaming for efficient data transfer to the data warehouse.

IMPLEMENTATION

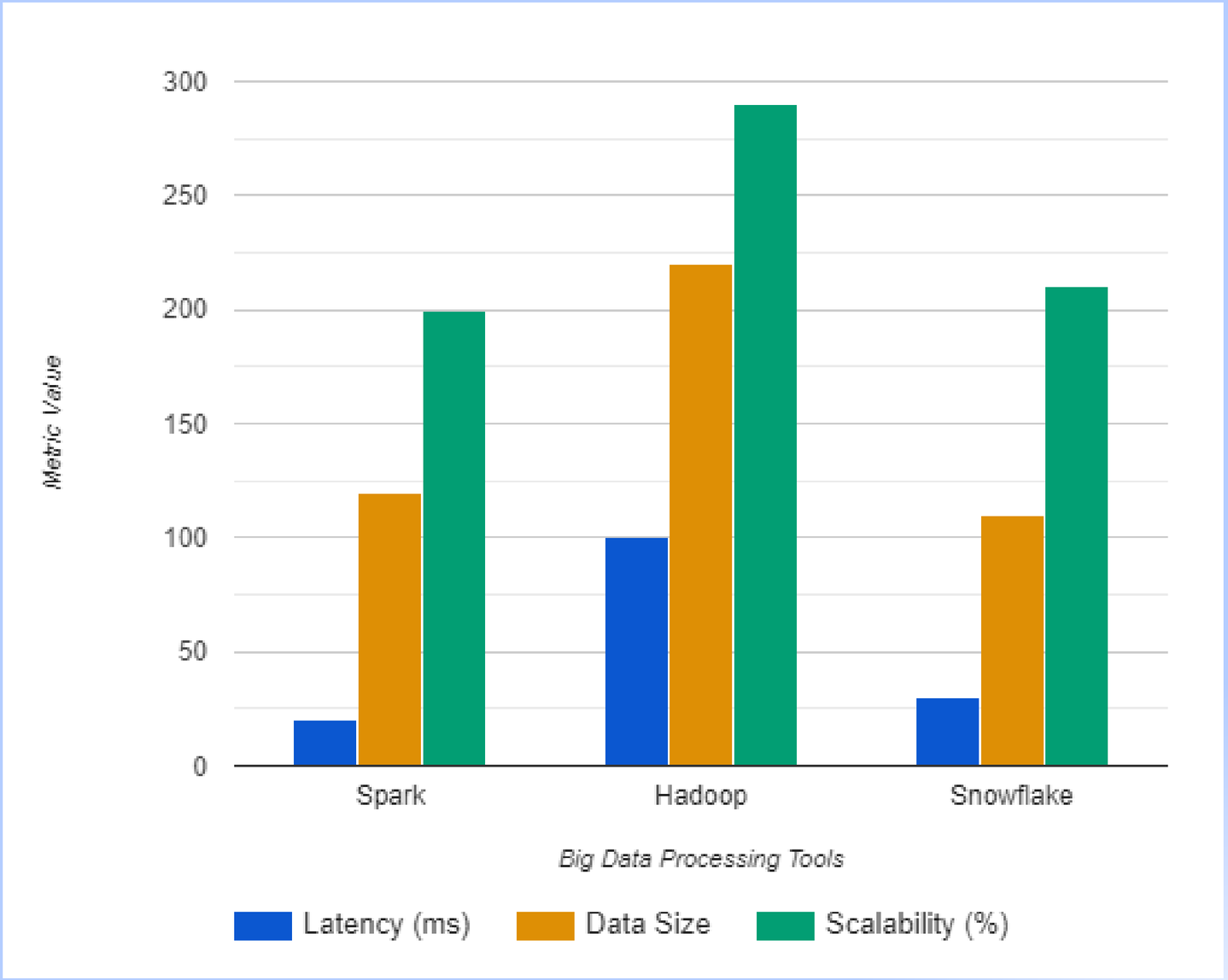
- Designing data warehousing architecture using Apache Spark for processing and analysis.

- Implementing strong encryption mechanisms for data security at both local and centralized levels.

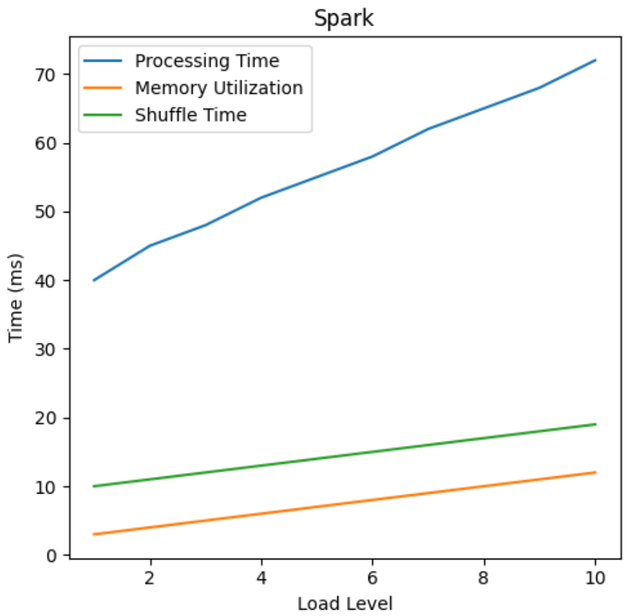
- Providing token-based and session-based authentication for user access control.

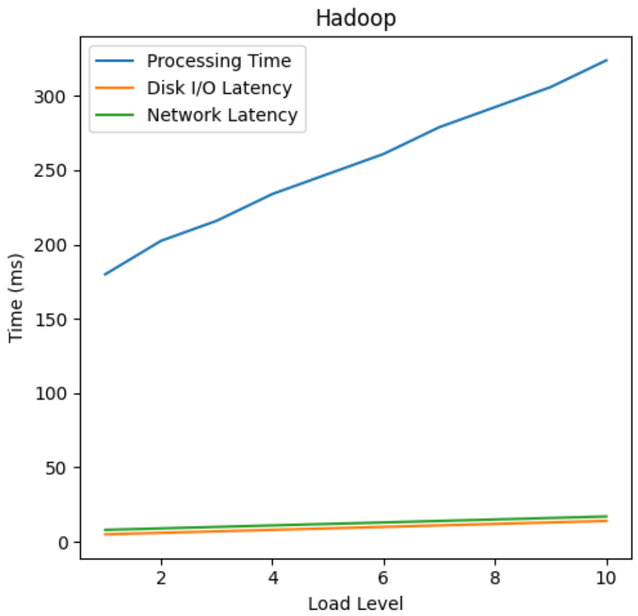
RESULTS

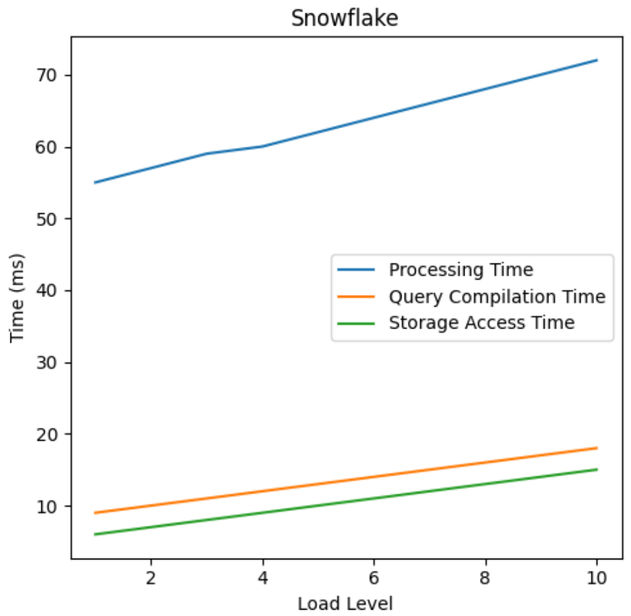
- Spark excels in scalability and efficient workload handling.
- It outperforms Snowflake in managing large datasets.
- Demonstrates superior resource utilization for cost-effectiveness.

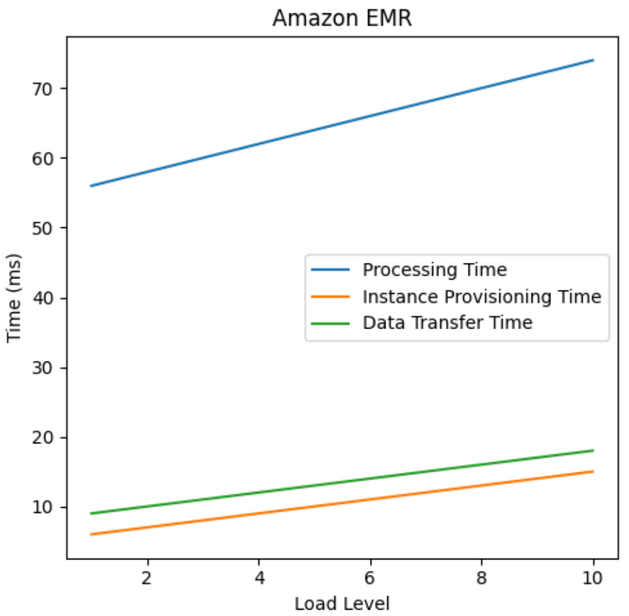


RESULTS









- Spark exhibits superior processing efficiency and resource utilization.
- Hadoop shows higher processing times and additional latencies.
- Snowflake and Amazon EMR demonstrate competitive processing times with slightly longer latencies.

Conclusions

The implementation of IPFS for decentralized storage and strong encryption mechanisms ensures enhanced security and privacy of electronic health records (EHR), safeguarding patient information from unauthorized access

Enhanced Security and Privacy

Leveraging Apache Spark for data warehousing enables efficient processing and analysis of EHR data, facilitating informed decision-making and improved patient care outcomes.

Efficient Data Warehousing

The developed framework offers scalability and interoperability, allowing seamless integration of EHR data from multiple branches and enabling comprehensive analytics across the entire healthcare organization.

Scalable and Interoperable Solution

TEAM CONTRIBUTION

VIJAI SURIA M

Designed and Developed the EHR Vault Application.
Drew the graphs and other visualizations
Documentation work

THANES M

Crafted the architectuure of EHR Vault
Analysed the past works
Performed Literature Survey
Documentation work

EHR VAULT

CONFIDENTIAL HEALTH REPOSITORY

THANK YOU!

REFERENCES

[1] M. M. Salim and J. H. Park, "Federated Learning-Based Secure Electronic Health Record Sharing Scheme in Medical Informatics," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, pp. 617-624, Feb. 2023, doi: 10.1109/JBHI.2022.3174823.

[2] A. F. Neamah, "Flexible Data Warehouse: Towards Building an Integrated Electronic Health Record Architecture," 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2020, pp. 1038-1042, doi: 10.1109/ICOSEC49089.2020.9215433.

[3] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE Journal of Biomedical and Health Informatics. 2018 Sep;22(5):1589-1604. doi: 10.1109/JBHI.2017.2767063.

[4] Z. M. Ibrahim et al., "A Knowledge Distillation Ensemble Framework for Predicting Short- and Long-Term Hospitalization Outcomes From Electronic Health Records Data," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 1, pp. 423-435, Jan. 2022, doi: 10.1109/JBHI.2021.3089287.

REFERENCES

[5] J. Li, X. Tan, X. Xu and F. Wang, "Efficient Mining Template of Predictive Temporal Clinical Event Patterns From Patient Electronic Medical Records," in IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 5, pp. 2138-2147, Sept. 2019, doi: 10.1109/JBHI.2018.2877255.

[6] A. Bahga and V. K. Madisetti, "A Cloud-based Approach for Interoperable Electronic Health Records (EHRs)," in IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 5, pp. 894-906, Sept. 2018, doi: 10.1109/JBHI.2013.2257818.

[7] Y. Li et al., "Hi-BEHRT: Hierarchical Transformer-Based Model for Accurate Prediction of Clinical Events Using Multimodal Longitudinal Electronic Health Records," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, pp. 1106-1117, Feb. 2023, doi: 10.1109/JBHI.2022.3224727.

[8] Y. Wang, P. -F. Li, Y. Tian, J. -J. Ren and J. -S. Li, "A Shared Decision-Making System for Diabetes Medication Choice Utilizing Electronic Health Record Data," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 5, pp. 1280-1287, Sept. 2017, doi: 10.1109/JBHI.2016.2614991.

REFERENCES

[9] Z. Huang, W. Dong, H. Duan and J. Liu, "A Regularized Deep Learning Approach for Clinical Risk Prediction of Acute Coronary Syndrome Using Electronic Health Records," in IEEE Transactions on Biomedical Engineering, vol. 65, no. 5, pp. 956-968, May 2018, doi: 10.1109/TBME.2017.2731158.

[10] Meng Y, Speier W, Ong M, Arnold CW. HCET: Hierarchical Clinical Embedding With Topic Modeling on Electronic Health Records for Predicting Future Depression. IEEE Journal of Biomedical and Health Informatics. 2021 Apr;25(4):1265-1272. doi: 10.1109/JBHI.2020.3004072.