Health Leadership and Quality of Life. 2024; 3:.585
doi: 10.56294/hl2024.585

HEALTH LEADERSHIP
AND QUALITY OF LIFE

ORIGINAL

Deep Learning-Based Natural Language Processing for the Identification and Multi-Label Categorization of Social Factors of Healthcare from Unorganized Electronic Medical Records

Procesamiento del lenguaje natural basado en Deep Learning para la identificación y categorización multietiqueta de factores sociales de la atención sanitaria a partir de historias clínicas electrónicas desorganizadas

Salim Davlatov[1], Isroil Sharipov[2], Dilrabo Mamatkulova[3], Dilnoza Boymatova[4], Mavsuma Oltiboyeva[5], Guzel Shamsutdinova[6], Natalya Kitayeva[7]

[1]Department of Faculty and Hospital Surgery, Bukhara State Medical Institute named after Abu Ali ibn Sino. Bukhara, Uzbekistan.
[2]Department of Anesthesiology, Resuscitation, and Emergency Medicine, Samarkand State Medical University. Uzbekistan.
[3]Department of Pediatrics No_3, Samarkand State Medical University. Uzbekistan.
[4]Department of Uzbek language and literature, Jizzakh state Pedagogical University. Uzbekistan.
[5]Department of Pharmaceutical Work Organization, Samarkand State Medical University. Samarkand, Uzbekistan.
[6]Department of Therapeutic Sciences, Fergana Medical Institute of Public Health. Uzbekistan.
[7]Department of disciplines in field of therapy, Fergana Medical Instutite of Public Health.

Cite as: G Davlatov S, Sharipov I, Mamatkulova D, Boymatova D, Oltiboyeva M, Shamsutdinova G, et al. Deep Learning-Based Natural Language Processing for the Identification and Multi-Label Categorization of Social Factors of Healthcare from Unorganized Electronic Medical Records. Health Leadership and Quality of Life. 2024; 3:.585. https://doi.org/10.56294/hl2024.585

Submitted: 01-05-2024 **Revised:** 25-09-2024 **Accepted:** 29-11-2024 **Published:** 30-11-2024

Editor: PhD. Prof. Neela Satheesh

Corresponding Author: Salim Davlatov

ABSTRACT

Social Factors of Healthcare (SFH) are non-medical determinants that may significantly influence patient health outcomes. Nevertheless, SFH is seldom included in Unorganized Electronic Medical Records (UEMR) data, such as diagnostic codes, and is often found in uncontrolled descriptive medical notes. Consequently, discerning social factors from UEMR data has gained paramount significance. Previous research towards using Natural Language Processing (NLP) for the automated extraction of SFH from text often emphasizes a selective approach to SFH. It fails to include the current advancements in Deep Learning (DL). This study proposes Deep Learning-Based Natural Language Processing for the identification and multi-label categorization (DL-NLP-MLC) of SFH from UEMR. Information was obtained from the Medical Information Mart for Intensive Care (MIMIC-III) dataset. The database consisted of 4 124 socially connected phrases derived from 2,785 medical notes. A framework for automatic MLC for multiple SFH types has been established. The database consisted of descriptive medical notes categorized as “SFH” inside the MIMIC-III medical dataset. Four types of categorization models have been trained: Decision Tree (DT), Random Forest (RF), and Long Short-Term Memory (LSTM). The efficacy of DL-NLP-MLC has been evaluated using accuracy, precision, recall, Area Under the Curve (AUC), and F1 score. The findings indicated that, in general, LSTM surpassed the other models of categorization with AUC (98,4 %) and Accuracy (94,6 %) for drug abuse SFH. The suggested method of training a DL classifier on a dataset rich in structured feature hierarchies may yield a very effective classifier using UEMR. Evidence demonstrates that model performance correlates with the semantic variety used by health practitioners and the automated creation of medical statements for documenting SFH.

Keywords: Social Factors; Healthcare; Deep Learning; Natural Language Processing; Multi-Label Categorization; Electronic Medical Records; Bidirectional Long Short-Term Memory.

RESUMEN

Los factores sociales de la atención sanitaria (SFH) son determinantes no médicos que pueden influir significativamente en los resultados de salud de los pacientes. Sin embargo, rara vez se incluyen en los datos de las historias clínicas electrónicas no organizadas, como los códigos de diagnóstico, y a menudo se encuentran en notas médicas descriptivas no controladas. En consecuencia, discernir los factores sociales a partir de los datos UEMR ha adquirido una importancia primordial. Las investigaciones anteriores sobre el uso del Procesamiento del Lenguaje Natural (PLN) para la extracción automatizada de la SFH a partir de texto suelen hacer hincapié en un enfoque selectivo de la SFH. No incluye los avances actuales en Deep Learning (DL). Este estudio propone el Procesamiento del Lenguaje Natural basado en el Aprendizaje Profundo para la identificación y categorización multietiqueta (DL-NLP-MLC) de SFH a partir de UEMR. La información se obtuvo del conjunto de datos Medical Information Mart for Intensive Care (MIMIC-III). La base de datos constaba de 4 124 frases socialmente conectadas derivadas de 2 785 notas médicas. Se ha establecido un marco para la MLC automática para múltiples tipos de SFH. La base de datos consistía en notas médicas descriptivas categorizadas como «SFH» dentro del conjunto de datos médicos MIMIC-III. Se han entrenado cuatro tipos de modelos de categorización: Árbol de decisión (DT), Bosque aleatorio (RF) y Memoria larga a corto plazo (LSTM). La eficacia de DL-NLP-MLC se ha evaluado mediante la exactitud, la precisión, la recuperación, el área bajo la curva (AUC) y la puntuación F1. Los resultados indicaron que, en general, LSTM superó a los otros modelos de categorización con AUC (98,4 %) y precisión (94,6 %) para SFH de abuso de drogas. El método sugerido de entrenar un clasificador DL en un conjunto de datos rico en jerarquías de características estructuradas puede producir un clasificador muy eficaz utilizando UEMR. Las pruebas demuestran que el rendimiento del modelo se correlaciona con la variedad semántica utilizada por los profesionales sanitarios y la creación automatizada de declaraciones médicas para documentar los SFH.

Palabras clave: Factores Sociales; Sanidad; Aprendizaje Profundo; Procesamiento del Lenguaje Natural; Categorización Multietiqueta; Historias Clínicas Electrónicas; Memoria Bidireccional a Corto Plazo a Largo Plazo.

INTRODUCTION

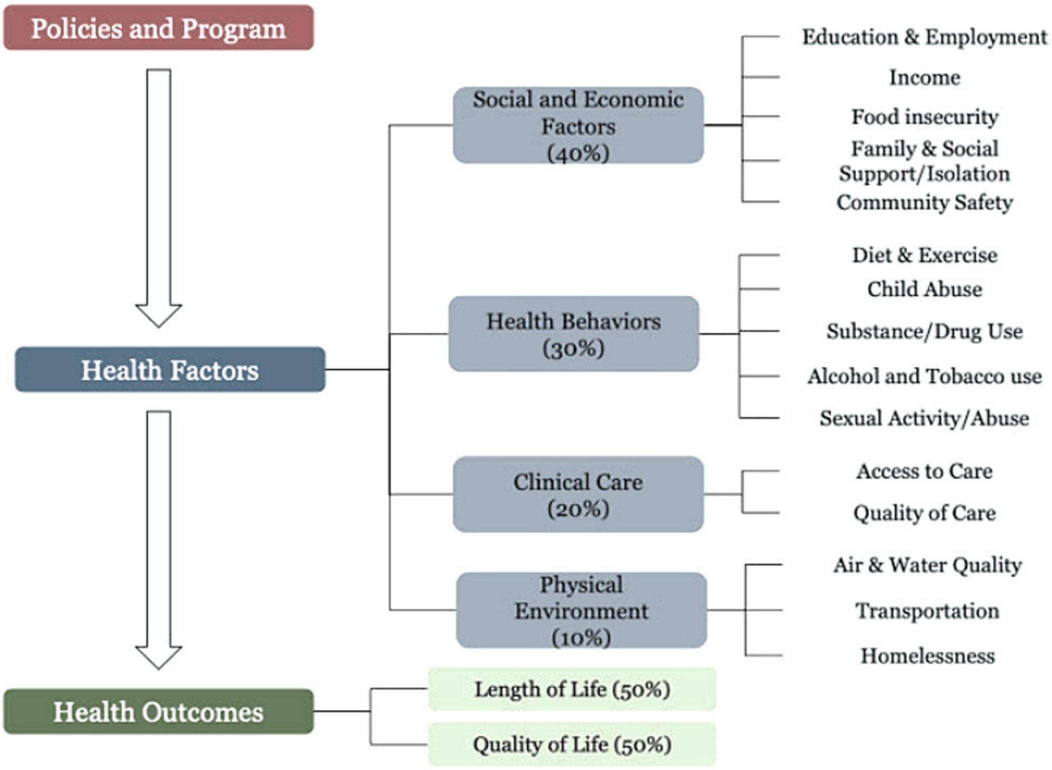

Figure 1. Enhancements in upstream initiatives and policies that influence SFH

SFH are conditions in which individuals are born, reside, acquire education, engage in employment, and age, and are intricately linked to their own medical behaviors, lifestyle choices, and social relationships. The allocation of wealth, authority, and assets at regional, national, and international levels may influence personal health conditions and possibly result in health inequalities. Numerous studies have examined the

https://doi.org/10.56294/hl2024.585

correlations between SFH and various medical conditions, including the impact of food insecurity on diabetes development[1] the influence of economic status, surroundings, jobs, race, and social assistance on the risk of breast cancer and longevity[2] and the impact of decent housing on psychological well-being.[3] Unsurprisingly, hazardous health behaviors and the inequitable distribution of SFH have been linked to heightened financial costs for both patients and doctors.[4]

The yearly medical rankings assess the influence of several health determinants by rating the wellness outcomes of more than 3,000 areas in the United States. Figure 1 suggests that enhancements in upstream initiatives and policies influence health determinants, affecting downstream public medical conditions.[5] Social, economic, and physical environment factors account for most health effects (53 %), while lifestyle choices contribute 32 %. Only 19 % of medical results are ascribed to clinical treatment. For instance, the Centers for Illness Control and Prevention (CIC) reports that 41 % of fatalities from persistent lower breathing illnesses are attributable to social and natural exposure to passive smoking, pollutants, workplace agents, and other indoor and outdoor air contaminants. To a lesser extent, 34 % of early stroke fatalities were ascribed to hazardous health habits (tobacco use, consumption of alcohol, and inactive lifestyles) and their associated clinical manifestations—hypertension, hypercholesterolemia, cardiovascular disease, diabetes, obesity, and prior stroke incidents.

This study indicates a significant correlation between nonclinical characteristics and clinical outcomes, enhancing medical and societal interest in integrating SFH into patient records. Gathering and comprehending SFH information has considerable promise and may provide critical contextual insights into patients' behaviors to enhance clinical results.[6] Most US healthcare institutions and practitioners use EMRs to record patient clinical data. Over the last ten years, the usage of EMRs has significantly increased; nonetheless, qualitative data about patients' lives is often recorded in unorganized medical notes. Frequently collecting SFH information is hindered by the absence of defined data items, evaluation methods, quantifiable inputs, and consistent data-gathering techniques in clinical notes, significantly restricting access to this data.

Traditionally, the extraction of meaningful information from unorganized information is done manually via chart inspection, a process that may be labor-intensive. Recent advancements in natural language processing provide more efficient, automated methods to extract and evaluate valuable information from current electronic medical record data. Presently, approximately 82 % of healthcare data is unorganized and does not conform to readily enforceable categories, such as doctor experience notes, summary of discharge, patient-reported data, and radiology/pathology documents; however, this data can be defined and utilized to facilitate better clinical choice-making.[7] Numerous clinical decision-support tools are being created utilizing electronic medical records, with numerous medical facilities dedicating significant resources to facilitate the integration of natural language processing technologies to augment the volume of usable information, enhance analytical insights, and enhance patient experiences.[8,9]

Recent research delineated the breadth and significance of NLP and data retrieval approaches for extracting SFH data from medical records.[10] Nonetheless, several different versions of SFH identification or extraction techniques exist in recent research, and the effectiveness of each tool is mostly contingent upon the specific form of SFH being addressed. Moreover, the absence of a thorough assessment that specifies the existing tools and their optimal applications may impede effective advancement on this research issue regarding the understanding of what has and has not been investigated. This paper examines several NLP strategies for curating the SFH vocabulary and developing SFH retrieval systems designed to extract SFH information from UEMR systems. It has enumerated EMR systems and classes of SFH ideas derived from NLP methods and technologies.

Chen et al.[11] examined the incorporation of SFH into EMRs, its effects on risk forecasting, and the resultant results. Agnikula et al.[12] investigated AI techniques for extracting SFH from EMRs. They briefly discussed various NLP techniques used to detect SFH from EMRs and reviewed the literature on medical results associated with SFH. Hybrid methodologies integrating NLP and ML represent the predominant biomedical strategy for retrieving clinical literature.[13] Numerous research successfully used NLP methodologies, including data extraction methods, for diverse classifications of SFH involving homelessness, work status, and dealing with violence.[14,15]

These approaches included routine patterns, named entity recognition (NER), and distributional semantics. Individuals experiencing socioeconomic hardship, such as job loss (employment instability), sometimes encounter many challenges related to this loss, including the termination of health insurance linked to their work. Moreover, text data in the healthcare industry is distinguished by lengthy phrases, including several technical terms and typographical errors.[16,17]

Innovative methodologies in machine learning, such as MLC, may be a suitable option for modeling the profiles of patients impacted by multiple SFH. MLC distinguishes itself from traditional ML by approaching the learning issue from an alternative angle. Unlike classical classification problems, where each observation is assigned to a single, mutually exclusive class, MLC's decision regions of labels (i.e., classes) intersect.

Binary significance, a conventional method for addressing the MLC challenge, disaggregates the issue into many autonomous binary categorization jobs, one for each label. Although research has been conducted on each specific SFH feature in the proposed model, it has been asserted that this work pioneers in addressing multi-labeling within the medical domain.

METHOD

This section delineates the proposed approach for precise, autonomous classification of SFH classes from UEMR data. The structure comprises an open dataset, tagged dataset phrases with SFH tags, sentence preparation procedures, and ML and DL modeling frameworks.

Dataset

The MIMIC-III medical dataset is a relational database encompassing extensive medical information about several patients admitted to the Intensive Care Unit (ICU). The collection of data(18) is publicly accessible and provided as either comma-separated integer formats or a singular Postgres database restore file. MIMIC-III v1.4 has been employed, including over 60 000 hospitalizations for 39 515 adults and 7 778 newborns. This anonymous dataset comprises extensive data about the clinical management of patients. The database included 4 124 socially interconnected terms extracted from 2 785 medical notes.

Preprocessing

Initially comprising 13 labeling classes (11 SFH categories, one substance abuse class, and one non-SFH class), this has been streamlined into eight distinct groups by retaining the seven most prevalent classes and consolidating the remaining labels into a singular "other" class due to their low frequency.

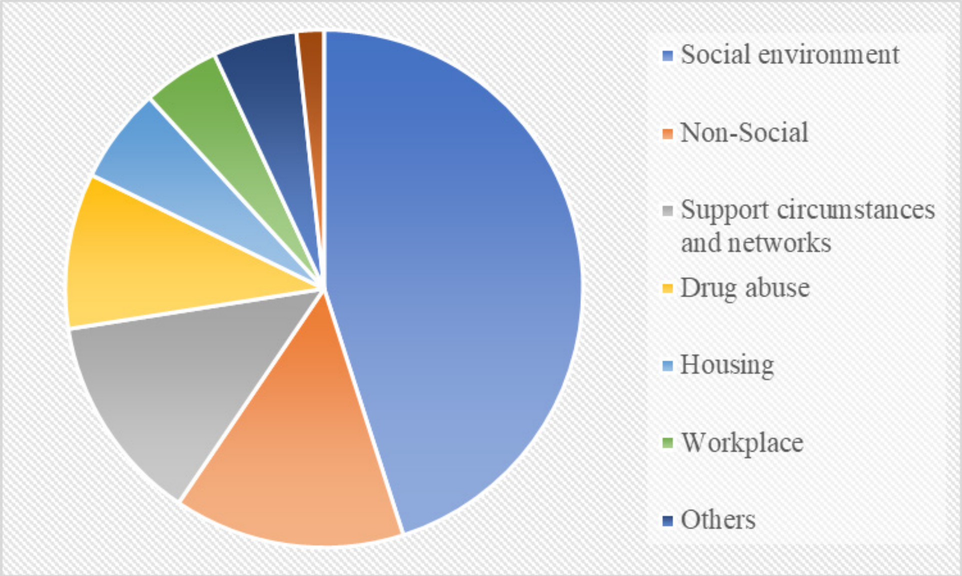

Figure 2. The conclusive pattern of annotation class prevalence

The conclusive pattern of annotation class prevalence is shown in figure 2. Despite consolidating the least common seven types into one, the resulting eight categories remained significantly uneven. Typical text processing has been performed on the 4 124 statements: the entire sentence has been converted to lowercase form, foreign or rare symbols were eliminated, and abbreviations were disaggregated.

DL-NLP-MLC of SFH from UEMR

Four categorization frameworks have been developed: Decision Tree (DT), Random Forest (RF), and Long Short-Term Memory (LSTM). We first acquired fundamental models like DT and RF using the training dataset. After the training stage, the model's efficacy was assessed by examining its estimations using the test information.

Decision Tree (DT): It is a non-parametric supervised learning method that partitions the input space into segments and assigns a class label to each segment.(19)

Random Forest (RF) is an ensemble learning method that improves predictive accuracy by integrating many decision trees (DTs).(20)

LSTM

Recurrent neural networks (RNNs) include the concept of time series in their architecture, making them particularly adept at analyzing temporal data. Conventional RNNs, however, encounter the problem of vanishing

https://doi.org/10.56294/hl2024.585

gradients as the temporal duration increases. LSTM is proposed as a solution to the long-dependence problem associated with RNNs. The LSTM employs a threshold mechanism to control the data acquisition rate and may choose to disregard previously retained information.

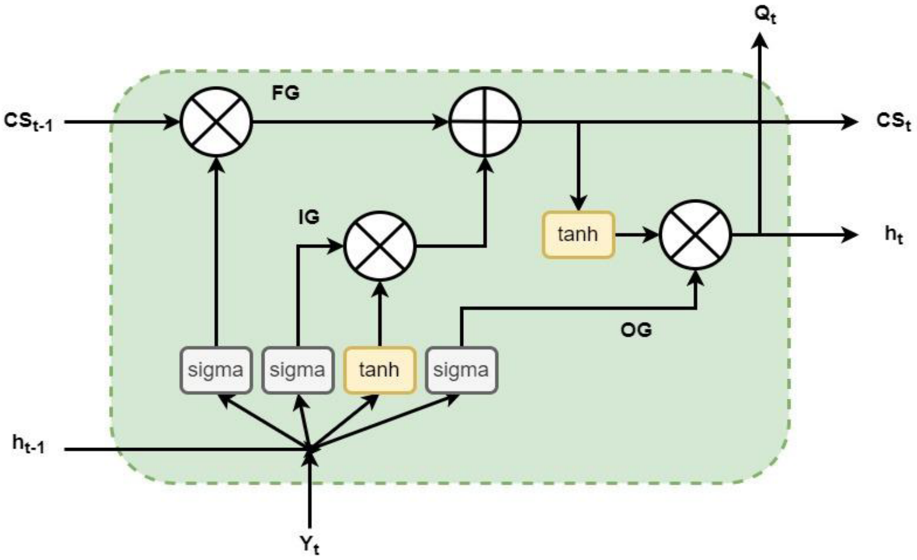

Figure 3. Basic structure of LSTM cell

LSTMs, akin to RNNs in basic architecture, possess a more intricate configuration for hidden layer components, as seen in figure 3. An Input Gate (IG), an Output Gate (OG), a Forget Gate (FG), and a storage cell are the components of a neuron. These gates circumvent the vanishing gradients problem by randomly determining the passage of data and its transmission to the cell. The IG specifies the data input to the cell, the OG provides the output from the cell, and the FG specifies the data eliminated from the cell. These three gates' opening and shutting hours are obtained from the network. A set of equations that elucidate the LSTM network's operation and modify its components' parameters has been given below.

Input Gate: It produces a value between 0 and 1, indicating the significance of the new data, and incorporates the current input value (Y_t) and the previous hidden state (h_{t-1}) is given in equation (1).

$$IG_t = \sigma(W_{IG}Y_t + W_{IG}h_{t-1} + b_{IG}) \qquad (1)$$

Forget Gate: It produces a value between 0 and 1, indicating the proportion of data to be discarded. Forget gate is given in equation (2).

$$FG_t = \sigma(W_{FG}Y_t + W_{FG}h_{t-1} + b_{FG}) \qquad (2)$$

Cell State (CS_t): It is altered according to the projected cell state (g_t), FG, and IG. Cell State is given in equation (3).

$$CS_t = (FG_t * CS_{t-1}) + (IG_t * g_t) \qquad (3)$$

The Output Gate (OG_t): It regulates the extent to which the current state of the cell is manifested as the output. Output gate is given in equation (4).

$$OG_t = \sigma(W_{OG}Y_t + W_{OG}h_{t-1} + b_{OG}) \qquad (4)$$

W and b in these equations represent the respective weights and biases of the LSTM network. The hyperbolic tangent function (tanh) constrains input values to the range of -1 to 1, whereas the sigmoid function (σ) limits input values to the interval of 0 to 1.

The MLC had been trained to employ the Decision Tree Classifier alongside the Random Forest Classifier from

Python's Scikit-Learn ML library.[21] To concurrently choose and assess models while preventing excessive fitting, layered cross-validating, including an exterior cross-validation cycle for model evaluation and an internal cross-validation cycle for model selection. In a single iteration of the 10-fold external cross-validation loop, one fold functions as a testing set for evaluating a chosen model, while the other nine folds are employed as a training set for model selection.[22] The learning set underwent a 5-fold inner cross-validation cycle using random searching (RS) for hyperparameter optimization since RS is often more effective than grid-based searches. A model with a specific hyperparameter pair has been trained on four folds and assessed on the fifth fold. The model, optimized with certain hyperparameters and exhibiting superior performance across all five folds, was rebuilt on the whole learning set and assessed on the testing set of the exterior cross-validation loop.[23] This procedure is carried out for all ten external folds.

NLP for SFH detection

The next approach pertains to the NLP algorithms used to identify SFH.[24] In contrast to certain studies for MLC involving five SFH, which developed distinct models to forecast individual SFH despite the possibility of multiple SFH within one note, we approached SFH identification as an MLC forecasting challenge. This enabled us to construct a singular model (for a specified ML or DL architecture) capable of simultaneously predicting multiple SFH classes. A singular model capable of predicting numerous classes may be more effective in clinical applications than using distinct models for each class.[25] We hypothesize that since all sentences, irrespective of meaning, require analogous (implicit) language processing (e.g., semantic interpreting or function labeling) within an LSTM for precise categorization, it is advantageous to amalgamate the training signals from various SFH labels into the training parameters of a singular model. An MLC arrangement facilitates transfer learning across multiple SFH categories.[26] Consequently, we hypothesize that a singular multi-label model may provide superior performance compared to individual models in every class; nevertheless, this hypothesis requires empirical validation.

Secondly, as mentioned in the introduction, we assessed a broader spectrum of NLP methodologies for SFH categorization than in other studies, including LSTM, traditional ML using bag-of-words, and three networks: DT, RF, and LSTM. The effectiveness of the models in this test, similar to other NLP tasks, mostly reflects the order seen in additional tasks: LSTM employs DL on bags of words.[27] We contend that this performance trend may be interpreted with considerable accuracy. Although the limited criteria for extracting SFH are precise, their scarcity results in high accuracy but poor recall of SFH. Bags of words yielded superior performance since they enable models to acquire a broader spectrum of separate words that forecast SFH or, in the context of nonlinear frameworks such as RF, a variety of words.[28]

Nonetheless, LSTMs surpassed bags of words due to (a) their use of previously trained GloVe embedded words using transfer learning, which is crucial in scenarios with little data, and (b) their integration of additional details on word structure and grammar.[29] Initial work, referenced in the subsequent section, proves that keyword and syllable identification may partly address the objective since specific words and phrases consistently trigger certain SFH classes. For instance, 'drug abuse' was often and consistently indicated by terms such as 'cocaine' or the phrase 'drug abuse' itself.

Ultimately, similar to other fields, we hypothesize that LSTM surpassed other DL networks due to its automatic attention mechanism and more comprehensive pre-training, including both the word-embedded level and all succeeding levels of the network. Initial training and transferable learning are essential in the context of little information.[30] The bag-of-words model eliminates the spatiotemporal data contained in plain text that DL networks are intended to use. Nonetheless, the reasons for the suboptimal performance of their feedforward network remain ambiguous, but we hypothesize that it may be mostly attributed to the need to identify the appropriate hyperparameters.[31] The research clearly illustrates that, when used correctly, LSTM may surpass conventional ML methods in identifying SFH.

RESULTS AND DISCUSSION

The MLC had been trained to employ the DecisionTreeClassifier alongside the Random Forest Classifier from Python's Scikit-Learn ML library.[21] To concurrently choose and assess models while preventing excessive fitting, layered cross-validating, including an exterior cross-validation cycle for model evaluation and an internal cross-validation cycle for model selection.[32] In a single iteration of the 10-fold external cross-validation loop, one fold functions as a testing set for evaluating a chosen model, while the other nine folds are employed as a training set for model selection.[33] The learning set underwent a 5-fold inner cross-validation cycle using random searching (RS) for hyperparameter optimization since RS is often more effective than grid-based searches.[34] A model with a specific hyperparameter pair has been trained on four folds and assessed on the fifth fold. The model, optimized with certain hyperparameters and exhibiting superior performance across all five folds, was rebuilt on the whole learning set and assessed on the testing set of the exterior cross-validation loop. This procedure is carried out for all ten external folds.

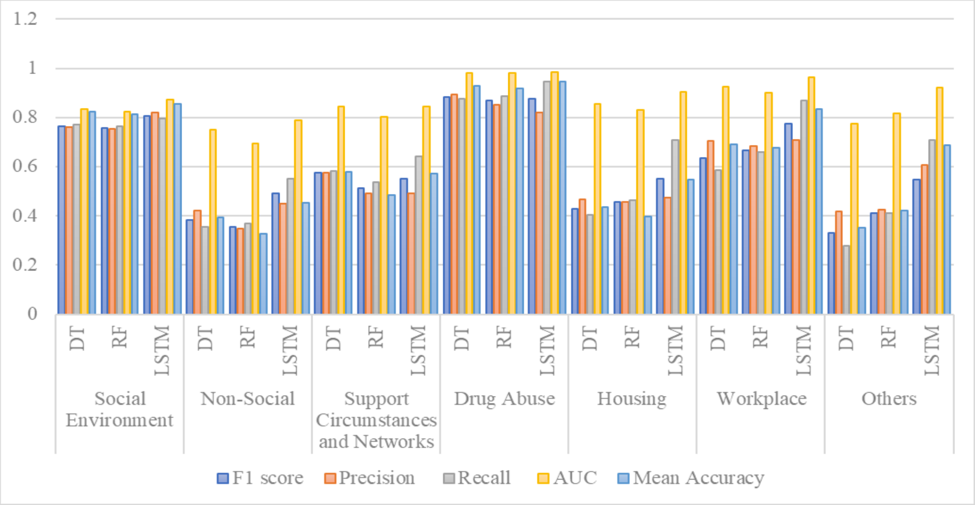

Figure 4. Effectiveness of three ML and DL models—DT, RF, and LSTM—across distinct annotation classes using numerous assessment metrics: F1 score, Precision, Recall, AUC, and Average Accuracy

Figure 4 delineates the effectiveness of three ML and DL models—DT, RF, and LSTM—across distinct annotation classes using numerous assessment metrics: F1 score, Precision, Recall, AUC, and Average Accuracy. BERT regularly surpasses other models across most classes, particularly in intricate domains such as Social Nature, Non-Social, and Supporting Circumstances, where it attains the greatest F1, Precision, and AUC values. In Drug Abuse and the Workplace, DT exhibits outstanding performance in F1 and Precision, whereas LSTM maintains robust overall efficacy. LSTM has exceptional performance in areas such as housing and others, particularly in recall and AUC, highlighting its efficacy in managing varied data while preserving high accuracy and recall. LSTM has exceptional performance, making it the optimal selection for this SFH categorization across most measures.

CONCLUSIONS

This paper presents Deep Learning-Based Natural Language Processing for the detection and multi-label classification of SFH from UEMR. Data were acquired from the MIMIC-III dataset. The database included 4,124 socially interconnected terms extracted from 2 785 medical notes. A framework for automated MLC for various SFH kinds has been developed. The database included descriptive medical notes classified as "SFH" inside the MIMIC-III medical dataset. Four categorization models have been developed: DT, RF, and LSTM. The results demonstrated that LSTM outperformed the other classification models with an AUC of 98,4 % and an accuracy of 94,6 % for drug abuse SFH. The proposed approach of training a deep learning classifier on a dataset rich in organized feature hierarchies may provide a very effective classifier using UEMR. Evidence indicates that model performance is correlated with the semantic diversity used by healthcare practitioners and the automated generation of medical documentation for SFH.

BIBLIOGRAPHIC REFERENCES

1. Robbiati, C., Armando, A., da Conceição, N., Putoto, G., & Cavallin, F. (2022). Association between diabetes and food insecurity in an urban setting in Angola: a case-control study. Scientific reports, 12(1), 1084.

2. Coughlin, S. S. (2021). Social determinants of health and cancer survivorship. Journal of environment and health sciences, 7(1), 11.

3. Marçal, K. (2024). Housing hardship and maternal mental health among renter households with young children. Psychiatry Research, 331, 115677.

4. Gadhia, S., Richards, G. C., Marriott, T., & Rose, J. (2023). Artificial intelligence and opioid use: a narrative review. BMJ Innovations, 9(2).

5. Magnan, S. (2021). Social determinants of health 201 for health care: Plan, do, study, act. NAM perspectives, 2021.

6. Truong, H. P., Luke, A. A., Hammond, G., Wadhera, R. K., Reidhead, M., & Maddox, K. E. J. (2020). Utilization of social determinants of health ICD-10 Z-codes among hospitalized patients in the United States, 2016-2017. Medical care, 58(12), 1037-1043.

7. Pramanik, M. I., Lau, R. Y., Azad, M. A. K., Hossain, M. S., Chowdhury, M. K. H., & Karmaker, B. K. (2020). Healthcare informatics and analytics in big data. Expert Systems with Applications, 152, 113388.

8. DeBarmore, B. M. (2022). Electronic Health Record Phenotyping in Cardiovascular Epidemiology (Doctoral dissertation, The University of North Carolina at Chapel Hill).

9. Reeves, R. M., Christensen, L., Brown, J. R., Conway, M., Levis, M., Gobbel, G. T., ... & Chapman, W. (2021). Adaptation of an NLP system to a new healthcare environment to identify social determinants of health. Journal of biomedical informatics, 120, 103851.

10. Hatef, E., Rouhizadeh, M., Tia, I., Lasser, E., Hill-Briggs, F., Marsteller, J., & Kharrazi, H. (2019). Assessing the availability of data on social and behavioral determinants in structured and unstructured electronic health records: a retrospective analysis of a multilevel health care system. JMIR medical informatics, 7(3), e13802.

11. Chen, M., Tan, X., & Padman, R. (2020). Social determinants of health in electronic health records and their impact on analysis and risk prediction: a systematic review. Journal of the American Medical Informatics Association, 27(11), 1764-1773.

12. Agnikula, K. S., & Balls-BerryJoyce Joy, E. (2021). Social and behavioral determinants of health in the era of artificial intelligence with electronic health records: a scoping review. Health Data Science.

13. Blosnich, J. R., Montgomery, A. E., Dichter, M. E., Gordon, A. J., Kavalieratos, D., Taylor, L., ... & Bossarte, R. M. (2020). Social determinants and military veterans' suicide ideation and attempt: a cross-sectional analysis of electronic health record data. Journal of general internal medicine, 35, 1759-1767.

14. Bettencourt-Silva, J. H., Mulligan, N., Sbodio, M., Segrave-Daly, J., Williams, R., Lopez, V., & Alzate, C. (2020). Discovering new social determinants of health concepts from unstructured data: framework and evaluation. In Digital Personalized Health and Medicine (pp. 173-177). IOS Press.

15. Topaz, M., Murga, L., Bar-Bachar, O., Cato, K., & Collins, S. (2019). Extracting alcohol and substance abuse status from clinical notes: The added value of nursing data. In MEDINFO 2019: Health and Well-being e-Networks for All (pp. 1056-1060). IOS Press.

16. Nock, M. K., Millner, A. J., Ross, E. L., Kennedy, C. J., Al-Suwaidi, M., Barak-Corren, Y., ... & Kessler, R. C. (2022). Prediction of suicide attempts using clinician assessment, patient self-report, and electronic health records. JAMA network open, 5(1), e2144373-e2144373.

17. Jose, T., Hays, J. T., & Warner, D. O. (2020). Improved documentation of electronic cigarette use in an electronic health record. International journal of environmental research and public health, 17(16), 5908. https://physionet.org/content/mimiciii/1.4/

18. Luo, H., Cheng, F., Yu, H., & Yi, Y. (2021). SDTR: Soft decision tree regressor for tabular data. IEEE Access, 9, 55999-56011.

19. Correia, A., Peharz, R., & de Campos, C. P. (2020). Joints in random forests. Advances in Neural Information Processing Systems, 33, 11404-11415.

20. Akusok, A., Leal, L. E., Björk, K. M., & Lendasse, A. (2021). Scikit-ELM: an extreme learning machine toolbox for dynamic and scalable learning. In Proceedings of ELM2019 9 (pp. 69-78). Springer International Publishing.

21. Grace Dolapo, P., Onanuga Ayotola, O., Ilori Olufemi, O., & Chukwuemeka Peter, U. (2020). Library Orientation and Information Literacy Skills as Correlates of Scholarly Research of Postgraduate Students of Federal University of Agriculture, Abeokuta, Nigeria. Indian Journal of Information Sources and Services, 10(1), 40–47. https://doi.org/10.51983/ijiss.2020.10.1.479

22. Knežević, D., & Knežević, N. (2019). Air Pollution-Present and Future Challenges, Case Study Sanitary Landfill Brijesnica in Bijeljina. Archives for Technical Sciences, 1(20), 73–80.

23. Konappa, D. (2020). Access and Use of Electronic Information Resources by Faculties of RGUKT, Andra Pradesh: A Study. Indian Journal of Information Sources and Services, 10(1), 7-12. https://doi.org/10.51983/ijiss.2020.10.1.484

24. Radmanović, S., Nikolić, N., & Đorđević, A. (2018). Humic Acids Optical Properties of Rendzina Soils in Diverse Environmental Conditions of Serbia. Archives for Technical Sciences, 1(18), 63-70.

25. Debbarma, K., & Praveen, K. (2019). LIS Education in India with the Emerging Trends in Libraries: Opportunities and Challenges. Indian Journal of Information Sources and Services, 9(S1), 41-43. https://doi.org/10.51983/ijiss.2019.9.S1.567

26. Cvijić, R., Milošević, A., Čelebić, M., & Kovačević, Ž. (2018). Geological and Economic Assessment of the Perspective of the Mining in Ljubija Ore Region. Archives for Technical Sciences, 1(18), 1-8.

27. Sobha Rani, J. (2019). A Study on Marketing Strategy for Library Resources and Services with Special Reference to Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh. Indian Journal of Information Sources and Services, 9(S1), 51-56. https://doi.org/10.51983/ijiss.2019.9.S1.564

28. Tunguz, V., Petrović, B., Malešević, Z., & Petronić, S. (2019). Soil and Radionuclides of Eastern Herzegovina. Archives for Technical Sciences, 1(20), 87-92.

29. Pal, F., & Hatua, S. R. (2019). Proficiency Building of Non-Academic Libraries in the Context of Present LIS Education in India: A Study. Indian Journal of Information Sources and Services, 9(1), 128-131. https://doi.org/10.51983/ijiss.2019.9.1.580

30. Karimov, A., et al. (2019). Rethinking settlements in arid environments: Case study from Uzbekistan. E3S Web of Conferences, 97, 05052. https://doi.org/10.1051/e3sconf/20199705052

31. Karimov, N., et al. (2024). Exploring food processing in natural science education: Practical applications and pedagogical techniques. Natural and Engineering Sciences, 9(2), 359-375. https://doi.org/10.28978/nesciences.1574453

32. Odilov, A., et al. (2024). Utilizing deep learning and the Internet of Things to monitor the health of aquatic ecosystems to conserve biodiversity. Natural and Engineering Sciences, 9(1), 72-83. https://doi.org/10.28978/nesciences.1491795

33. Balasundaram, A., Routray, S., Prabu, A. V., Krishnan, P., Malla, P. P., &amp; Maiti, M. (2023). Internet of Things (IoT)-based smart healthcare system for efficient diagnostics of health parameters of patients in emergency care. IEEE Internet of Things Journal, 10(21), 18563-18570.

34. Ebenezar, U. S., Vennila, G., Balakrishnan, T. S., &amp; Krishnan, P. (2024, June). Optimizing Healthcare Delivery through CloudBased Clinical Decision Support Systems. In 2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0 (pp. 1-6). IEEE.

FINANCING

None.

CONFLICT OF INTEREST

Authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Salim Davlatov, Isroil Sharipov, Dilrabo Mamatkulova, Dilnoza Boymatova, Mavsuma Oltiboyeva, Guzel Shamsutdinova, Natalya Kitayeva.

Data curation: Salim Davlatov, Isroil Sharipov, Dilrabo Mamatkulova, Dilnoza Boymatova, Mavsuma Oltiboyeva, Guzel Shamsutdinova, Natalya Kitayeva.

Formal analysis: Salim Davlatov, Isroil Sharipov, Dilrabo Mamatkulova, Dilnoza Boymatova, Mavsuma Oltiboyeva, Guzel Shamsutdinova, Natalya Kitayeva.

Drafting - original draft: Salim Davlatov, Isroil Sharipov, Dilrabo Mamatkulova, Dilnoza Boymatova, Mavsuma Oltiboyeva, Guzel Shamsutdinova, Natalya Kitayeva.

Writing - proofreading and editing: Salim Davlatov, Isroil Sharipov, Dilrabo Mamatkulova, Dilnoza Boymatova, Mavsuma Oltiboyeva, Guzel Shamsutdinova, Natalya Kitayeva.