Journal of Information Systems Engineering and Management
2025, 10(18s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Early Stage Disease Identification in a Context-aware Environment Leveraging Digital Security and Natural Language Processing

P. Karthikeyan[1], Dr. S. Geetha[2], S. Tharnidhar[3], R. Yuvaraj[4], S. Sriram[5], S. Jayaraj[6]
[1] Research Scholar, Department of Banking Technology, Pondicherry University, Puducherry, India.
[2] Assistant Professor, Department of Banking Technology, Pondicherry University, Puducherry, India.
[3] Department of Computer Science and Engineering, Sri Manakula Vinayagar Engineering College, Puducherry, India.
[4] Department of Computer Science and Engineering, Sri Manakula Vinayagar Engineering College, Puducherry, India.
[5] Department of Computer Science and Engineering, Sri Manakula Vinayagar Engineering College, Puducherry, India.
[6] Department of Computer Science and Engineering, Sri Manakula Vinayagar Engineering College, Puducherry, India.

ARTICLE INFO

Received: 15 Dec 2024

Revised: 26 Jan 2025

Accepted: 15 Feb 2025

ABSTRACT

Using natural language processing and machine learning, an innovative system for the diagnosis of early-stage diseases is developed within a context-aware environment. The system combines extracted symptoms with contextual data, such as age, gender, and location, to enhance the accuracy of diagnosis. For symptom-specific studies, using a reinforcement learning framework ensures better results in cases of uncertain or insufficient input. By putting together patient demographics and researching symptom development, this method will provide a more specific and precise diagnosis. The outcomes of health care should therefore be efficiently improved through this sophisticated AI-driven approach
Keywords: disease diagnosis, natural language processing, machine learning, context-aware systems, reinforcement learning, symptom extraction, healthcare AI, patient demographics, diagnostic accuracy, symptom progression, personalized medicine, early detection, medical informatics, AI in healthcare, symptom-specific investigation, digital security.

INTRODUCTION

Fast developments in artificial intelligence are revolutionizing almost all industries and, among all, the medical sector is among the most strongly affected. Classic disease diagnosis essentially depends on traditional medical skills. However, in the present increasingly complex scenario, AI-based solution is coming forth as a feasible tool for better diagnosis and cure. Essential for these AI-driven systems are the ML and NLP techniques involved, allowing extensive volumes of data to be parsed and consequently further increasing the accuracies of forecasts in real time. The capabilities of AI-powered diagnostic tools stand out, even in early diagnostics of disease development, especially considering shortcomings within general healthcare systems.

Recent study research has identified that ML algorithms can analyze EHRs and bring out patterns related to various diseases. To come up with exact predictions, the algorithms can check demographic data, medical histories, and symptoms a patient may have. But concerns exist regarding the precision and personalization of these predictions when symptoms seem vague or overlapping. These issues should be dealt with in a proper manner using context-awareness. Context-aware computing improves the correctness of diagnosis based on the user's surroundings that could include the location, gender, and age.

With RL embedded in the diagnostics framework, the system will learn and adapt as time progresses. Even while interacting with and learning from the symptoms during assessment, the feedback regarding the appropriate questions to be asked to reach an accurate, personalized diagnosis from this RL model, it will then make better-informed choices. When such symptoms are ambiguous and direct knowledge is not available for a specific part of medicine, this process is important. Reinforcement learning is proved to dramatically improve decision-making in related applications, therefore it is an important part of the suggested system.

Context awareness is another function in the detailing of a diagnosis of a disease. Patients' demographics differ in various ways-from age and gender to location and history, all affecting risk as well as the progression of the disease. The contextual data assimilated into the system helps in refining the predictions with a view relevantly to the patient's background. This serves to fill the gap between generalized models of diseases and personal care while at the same time targeting at improving the capacity of the diagnosis system.

The second important component is natural language processing. NLP makes the system to understand and also allows it to process human language. NLP simply means that the system can pull relevant symptoms from the patient's narrative input to map them properly. Advanced techniques such as NER and sentiment analysis enable the system to better identify the symptoms but also how emotionally distraught or urgent the patient is. That would consequently lead to the most accurate and actionable results.

In summary, the proposed system integrates multiple cutting-edge technologies—machine learning, reinforcement learning, and NLP—into a unified framework designed for early disease diagnosis. By leveraging context-aware capabilities and personalized symptom analysis, the system offers a promising solution to improve healthcare outcomes. The combination of these technologies promises not only to streamline the diagnostic process but also to enhance the overall patient experience, reducing the burden on healthcare professionals while providing timely and accurate diagnoses.

LITERATURE REVIEW

ML applications in healthcare were a huge subject of attention because it promised the betterment of diagnostics and results of patients. Particularly, the algorithms applied on large databases of patient records, lab data, and demographical information were used to predict the disease results and early diagnostic patterns. In medicine, these methods have been useful in many sectors, especially those involving oncology, cardiology, and neurology, because early detection makes a big difference. It is found that by training ML on huge, high-quality datasets, the models may predict the prognosis of several diseases with accuracy to the finest granularity by detecting some intricate relationships amongst the data. [4].

Context-aware systems with consideration of demographic and environmental factors in the determination of disease is also important. The healthcare systems will offer greater personalization and accuracy by the incorporation of these context-aware factors like age, gender, and location that might influence disease risks in a very large scale. According to Zou et al. [5], these systems support the context-aware framework in medical diagnosis, since such systems aid in refining the prediction because they have a better view of a patient's condition. These systems can change their behavior dynamically, due to context information provided by the patient himself or herself, in uncertain situations plays a key role in taking a better decision.

The usefulness of RL in the medical diagnosis of cases is also being established. This algorithm is to learn from its outcomes and use the knowledge as an optimization approach to decision making over time. Its adaptive mechanism can be applied towards improvement in the healthcare system and in areas like human judgment for symptom diagnosis. Tiwari et al. [6] designed a dialogue agent for disease diagnosis where a hierarchical reinforcement learning led the system in an iterative query of symptoms in order to process ambiguous and incomplete symptom inputs. The RL framework improved the decision-making process of the system, learning from previous interactions, in order to increase the accuracy and personalization of the diagnosis.

Thus, through NLP, the unstructured data of patients- in the form of medical records, clinician notes, and narratives-overcome to produce meaningful insights. NLP essentially becomes an indispensable step in extracting symptoms, conditions, and other relevant information from textual inputs and helps an AI system to grasp a patient's concern easily. Significant research has been conducted on NLP and its potentiality within the health domain. Such may make the proposed diagnostic models effective. This is proved by the research study of Patil et al. [7]. In that, it is evident that the combination of CNNs with NLP could make the evaluation of patient narratives combined with medical history to provide an early prediction for Alzheimer's disease.

NLP and ML have shown to be efficient in extracting symptoms and predicting diseases, but still, it remains an area of difficulty due to heterogeneity in the patient data. Variability in symptom presentation and the presence of incomplete or noisy data poses an obstacle to prediction accuracy, hence the need for data preprocessing techniques such as normalization and missing data imputation as discussed by Ramadhan et al. [8]. Integration of machine learning, NLP, and reinforcement learning has a huge potential for early-stage disease diagnosis, but that is possible only by addressing factors such as data quality, model interpretability, and patient context. Recent advances in explainable AI (XAI) have allowed greater transparency in understanding model predictions, thus fostering trust and ensuring reliable applications in healthcare, as emphasized by Dharmarathne and Bogahawaththa [9]

Apart from these technological advances, ethical considerations relating to the development of artificial intelligence in healthcare include questions regarding privacy and security over patients' health information as well as potential issues with biased decisions in AI-supported diagnosis. Noting the need to place strong policies for safeguarding patient data, Azeez and Van der Vyver [10] provide an in-depth study of the security and privacy issues surrounding e-health systems. They stress the importance of open data-sharing systems and also highlight the role that blockchain technology plays in integrity and privacy within medical records. There must be an ethical response to the development of healthcare systems that include artificial intelligence in order to be able to build trust among patients and instill a sense of responsible use of technology.

EXISTING SYSTEM

Many of the disease diagnosis systems in place today are still founded on the experience of health practitioners and the use of traditional diagnostic methods and tools. Its reliance now is so much on the manual assessment of clinical data, which include results from laboratory tests and patient history. Most of these have resulted in errors or delayed detection of diseases. Medical professionals are able to analyze structured data but have difficulty in analyzing unstructured data, which might be a limitation in early-stage diagnosis. The proposed solution of overcoming this limitation is machine learning algorithms, offering a more data-driven approach for better accuracy and speed in diagnostics [10].

However, the existing systems do not include unstructured sources of data such as clinical notes, medical images, and patient narratives. NLP has emerged as a critical tool to bridge the gap. This is because the meaningful information drawn from free-text data can enhance the effectiveness of disease prediction models. Several researchers have proposed methods for leveraging NLP in conjunction with machine learning algorithms to enhance the performance of diagnostic systems, particularly when dealing with diverse and complex data sources [11].

The system also has medical diagnosis systems based on context-aware computing. Information in the demographics of a patient like age, gender, or location can easily help in a better and accurate personalized diagnosis. Having such knowledge related to environment or personal considerations would further reinforce knowing the status of the patient in order to deliver more correct predictions. Significantly, experiments highlighted several evidence-based factors proving diagnostic accuracy may indeed be better by considering some environmental and personal elements [12].

The next promising area of research in the development of disease diagnosis systems involves reinforcement learning. Traditional systems function on pre-defined rules. Reinforcement learning, however, permits systems to learn and improve with experience over time through feedback while in the diagnosis process. This is a kind of learning that changes dynamically based on diagnosis from time to time as more data is accrued. Through the simulation of different scenarios and refinement of decision-making strategies, RL-based systems can continue to optimize their diagnostic accuracy to make them ideal for dealing with complex and ever-evolving disease profiles [13].

Data quality is critical in the diagnostic system. Incomplete or noisy data may lead to incorrect predictions and early-stage detection of diseases that might have occurred. Effective preprocessing techniques, feature extraction, normalization, and imputation all contribute to high accuracy in models for machine learning, especially with chronic disease prediction [14]. However, since the models do not have sufficient interpretability, they are hardly accepted in the clinical practice domain. The XAI methods address this by allowing for transparent insights in decision-making. This facilitates the trustedness of and helps in validation of AI-based diagnoses by healthcare professionals [15].

ALGORITHM

In the proposed disease diagnosis system, several advanced algorithms will be implemented to optimize functionality, accuracy, and adaptability. The primary algorithms are as follows:

1. *Random Forest Classifier*

   The random forest classifier uses an ensemble of decision trees to make predictions. It aggregates the predictions of multiple decision trees to improve classification accuracy. The formula for a decision tree in the random forest is given by:

$$\tilde{y} = \arg\max_{r} \sum_{s=1}^{N} I(f_s(w) = r)$$

Where:

- $f_s(w)$ is the prediction of tree s^{th} decision tree
- I is an indicator function that takes the value 1 if the condition holds and 0 otherwise.
- r is the class label

2. *Natural Language Processing*

NLP techniques are used to process and extract relevant information from unstructured patient input, such as symptoms described in natural language. The primary tasks in NLP are text preprocessing, symptom extraction, and sentiment analysis

$$TF\text{-}IDF(a,b) = TF(a,b) \times IDF(a)$$

Where:

- *TF(a, b)* represents the term frequency of word a in document b, measuring how often it appears.
- *IDF(a)* is the inverse document frequency of the term a across all documents.
- *TF-IDF(a, b)*combines both values to rank the importance of a in b.

3. *Reinforcement Learning (RL) Algorithm*

Reinforcement learning is used to adapt the system based on patient feedback and interactions. The RL agent learns to select the most relevant symptoms for further investigation. The algorithm is driven by the reward function, which encourages the agent to ask questions that lead to better disease predictions. The fundamental formula for RL in this context is the Bellman equation:

$$Q(s,a) = R(s,a) + \gamma \max_{a'} Q(s',a')$$

Where:

- *Q(s, a)* is the action-value function representing the expected cumulative reward of taking action *a* in state *s*,
- *R(s, a)* is the immediate reward after taking action *a* in state *s*
- γ is the discount factor,
- $max_{a'}Q(s', a')$ is the maximum future reward for the next state *s′*.

4. *Context-Aware Prediction Algorithm*

Incorporating contextual data, such as patient demographics (age, gender, location) into the disease prediction model helps to refine the diagnosis. The algorithm used here combines classical machine learning with contextual input through feature engineering. A general formula for context-aware learning can be represented as:

$$\hat{y} = f(X, C)$$

Where:

- *y^* is the predicted disease outcome,
- *X* represents the symptom features,
- *C* represents the contextual features (age, gender, location),
- *f* is the predictive model trained to consider both *X* and *C*.

PROBLEM STATEMENT

The primary challenge in disease diagnosis lies in accurately identifying potential conditions from limited and unstructured symptom data provided by patients. Traditional diagnostic methods often fail to account for contextual factors such as age, gender, and location, leading to misdiagnoses or delays in treatment. Existing systems also lack adaptability and transparency, reducing their effectiveness and trustworthiness. This project aims to develop an intelligent, context-aware disease diagnosis system that leverages machine learning, reinforcement learning, and NLP to enhance accuracy and explainability. The system will provide personalized predictions, iterative symptom investigation, and actionable insights to improve healthcare.

PROPOSED SYSTEM

The proposed system introduces a novel approach to the diagnosis of disease by integrating the concepts of context-aware computing, reinforcement learning, and natural language processing. While traditional systems use only structured information, this architecture collects both unstructured and structured data, incorporating user demographics: age, gender, and location, as well as symptom narratives, to render personalized predictions. NLP techniques preprocess and abstract meaningful insights from user input. This feature occurs due to the more accurate translation of symptoms into potential diseases. In this sense, the system's adaptive ability regarding symptom progression or specific contexts improves the diagnostic accuracy. Also, the knowledge base is continually updated and user input forms feedback loops.

A symptom investigation module uses reinforcement learning-based mechanisms to alleviate ambiguities or lack of input symptom data coming from the user. The symptom investigation module would iteratively prompt the user with relevant symptoms via contextual knowledge, as well as a pre-trained symptom-disease association model, in order to ensure hierarchical learning and thus proper convergence of potential diseases on which the system has converged. By integrating user-provided data with contextual analysis, the proposed solution avoids the limitations of fixed-rule systems, thereby improving accuracy and user experience

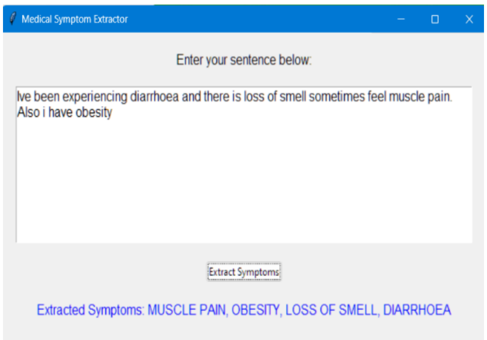


Figure 1. Symptom Investigation Module

The system then further makes disease predictions through a multi-modal learning framework that includes textual, contextual, and demographic data. Machine learning models further generate confidence scores for each of the predicted conditions to ensure transparency and reliability in the decision-making process. This further allows the system to explain its predictions, making the user more actionable in the sense that he or she knows whether he or she needs to visit a healthcare provider or take certain diagnostic tests, for instance. This design improves the whole diagnostic process, as well as catering to the needs of modern healthcare with personalization, explainability, and context awareness.

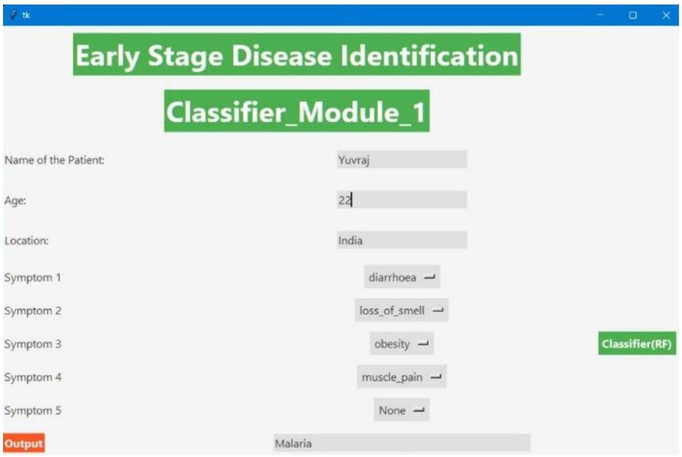


Figure 2. Symptom Classifier Module

Designed with the challenges of modern health care in mind, the system is aligned towards the principles of personalized medicine - context-aware, explainable, and artificial. Its continuously up-to-date knowledge base and its feedback loops give it the scope to adapt quickly to changing knowledge in medicine as well as for users. As a solution filling the gaps within traditional rule-based systems, the diagnostic precision alone is enhanced. It redefines the user experience and makes health care more accessible, intelligent, and user-centered.

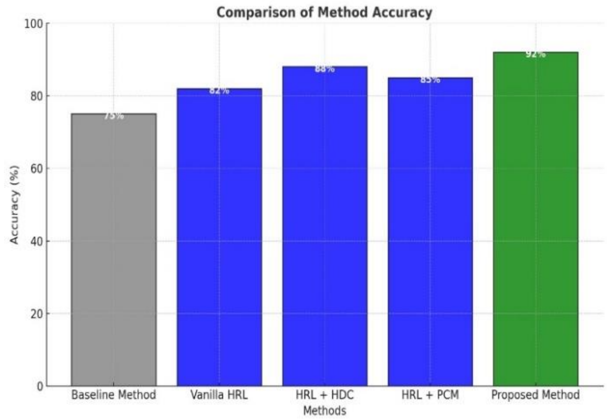
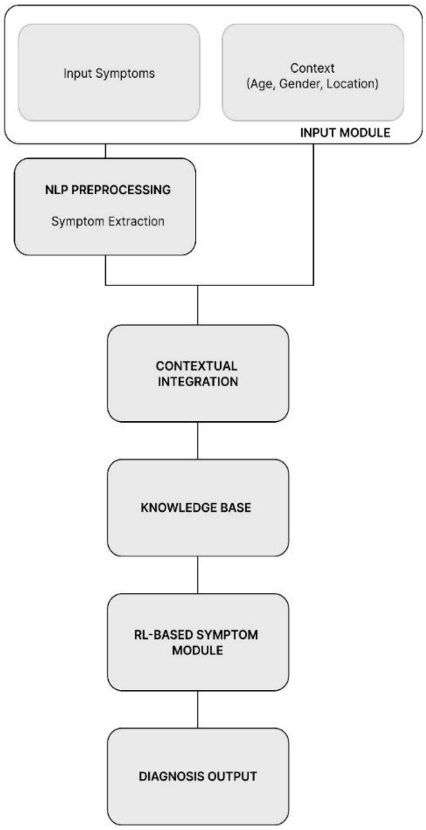


Figure 3. Model Comparison Analysis

ARCHITECTURE DIAGRAM OF THE PROJECT



CONCLUSION

In summary, the proposed system will address all of the issues found in the existing models for disease diagnosis. By utilizing hierarchical reinforcement learning, multi-modal data integration, and context-aware decision-making, the system is intended to produce accurate, personalized, and dynamic diagnostic results. Moreover, unstructured data can be assimilated using natural language processing for the better comprehension of complex patient narratives and symptoms, thereby enabling more informed and reliable diagnoses. This is holistic, thus it can work well with the complexities of the real world such as incomplete or ambiguous data.

The explainable AI techniques will thus address the major concerns of transparency in models; hence, this system will bring trust and facilitate the interpretation process for healthcare providers. The potential of the continuous adaptation and fine-tuning in diagnoses and treatments based on increasingly available patient data will improve healthcare delivery as a whole. The proposed system holds the promise to significantly enhance the accuracy of diagnostics, individualize treatment, and simplify health processes. Therefore, it stands at a huge leap in applying AI to health.

REFERENCES

[1] Shyam Visweswaran,Derek C. Angus , Margaret Hsieh, Lisa Weissfeld," Learning patient-specific predictive models from clinical data"
[2] Shujun Zou, Kaiyu Wan, Zongyuan Yang" Modelling and Verifying of Medical Diagnosis System Based on Context-awareness Framework"
[3] Abhisek Tiwari , Sriparna Saha, Pushpak Bhattacharyya" A knowledge infused context driven dialogue agent for disease diagnosis using hierarchical reinforcement learning"
[4] Gangani Dharmarathne , Madhusha Bogahawaththa" On the diagnosis of chronic kidney disease using a machine learning-based interface with explainable artificial intelligence"
[5] P Kumar V Subathra Y Swasthika; V Vishal "Disease Diagnosis Using Machine Learning on Electronic Health Records"
[6] Priyanka S; Keerthika S; Santhiya S; S. Jenesha Malar; A. Malavika; T. Dinesh Babu "Enhancing Disease Diagnosis with Machine Learning Based Symtom Prediction"
[7] Tashi Lhamo; Geeta Arora; Homan Emadifar; Ezekiel Olaoluwa Omole; Laith Abualigah; Peter Onu "Heart Disease Diagnosis by Machine Learning Techniques"
[8] Divyansh Nishad; Anshika Mishra; Nidhi Goyal "Symptom-Based Disease Prediction Using Machine Learning"
[9] Nur Ghaniaviyanto Ramadhan; Adiwijaya; Warih Maharani; Alfian Akbar Gozali "Chronic Diseases Prediction Using Machine Learning With Data Preprocessing Handling: A Critical Review"
[10] Mohammed Badawy, Nagy Ramadan & Hesham Ahmed Hefny "Healthcare predictive analytics using machine learning and deep learning techniques: a survey"
[11] P. Celard1,2,3 • E. L. Iglesias1,2,3 • J. M. Sorribes-Fdez1,2,3 • R. Romero1,2,3 • A. Seara Vieira1,2,3 • L. Borrajo1,2, A survey on deep learning applied to medical images: from simple artificial neural networks to generative models
[12] Vijeeta Patil1 , Manohar Madgi1 and Ajmeera Kiran Early prediction of Alzheimer's disease using convolutional neural network: a review
[13] Shahadat Uddin1 , Arif Khan1,2, Md Ekramul Hossain1 and Mohammad Ali Moni3Comparing different supervised machine learning algorithms for disease prediction
[14] Zhenxing Xu1 , Chang Su2 , Yunyu Xiao 1 , Fei WangArtificial intelligence for COVID-19: battling the pandemic with computational intelligence
[15] Chaitanya Singh a , Deepika Chauhan a , Sushama A. Deshmukh b , Swati Sudhakar Vishnu c , Ranjan Walia dMedi-Block record: Secure data sharing using block chain technology
[16] Nureni Ayofe Azeez, Charles Van der VyverSecurity and privacy issues in e-health cloud-based system: A comprehensive content analysis
[17] Mumtaz Ali1 , Muhammad Arshad1 , Ijaz Uddin1 , Muhammad Binsawad2 , Abdullah Bin Sawad3 and Osama Sohaib4Efficient context-aware computing: a systematic model for dynamic working memory updates in context-aware computing.
[18] K. Liao, Q. Liu, Z. Wei, B. Peng, Q. Chen, W. Sun, X. Huang, Taskoriented dialogue system for automatic disease diagnosis via hierarchical reinforcement learning, 2020, arXiv preprint.
[19] R.J. Byrd, S.R. Steinhubl, J. Sun, S. Ebadollahi, W.F. Stewart, Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records, Int. J. Med. Inform. 83 (2014) 983-992.
[20] Z. Wei, Q. Liu, B. Peng, H. Tou, T. Chen, X.-J. Huang, K.-F. Wong, X. Dai, Task-oriented dialogue system for automatic diagnosis, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Volume 2: Short Papers, 2018, pp. 201-207.
[21] X. Li, Y.-N. Chen, L. Li, J. Gao, A. Celikyilmaz, End-to-end task-completion neural dialogue systems, in: Proceedings of the Eighth International Joint Conference on Natural Language Processing, 1: Long Papers, 2017, pp. 733-743
[22] B. Liu, G. Tur, D. Hakkani-Tur, P. Shah, L. Heck, Dialogue learning with hu- man teaching and feedback in end-to-end trainable task-oriented dialogue systems, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, 2018, pp. 2060-2069.
[23] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, Nature 518 (2015) 529-533.
[24] A.G. Barto, S. Mahadevan, Recent advances in hierarchical reinforcement learning. Discrete Event Dyn. Syst, 13 (2003) 41-77.

[25] R. Parr, S. Russell, Reinforcement learning with hierarchies of machines, in: Advances in Neural Information Processing Systems, 1998, pp. 1043-1049
[26] L Baird, Residual algorithms: Reinforcement learning with function ap- proximation, in: Machine Learning Proceedings. Vol. 1995, 1995, pp. 30-37.
[27] R.S. Sutton, D. Precup, S. Singh, Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning, Artificial Intelligence 112 (1999) 181-211.
[28] K.-F. Tang, H.-C. Kao, C.-N. Chou, E.Y. Chang, Inquire and diagnose: Neural symptom checking ensemble using deep reinforcement learning, in: NIPS Workshop on Deep Reinforcement Learning, 2016.
[29] Liu, F. Pan, L. Luo, Gochat: Goal-oriented chatbots with hierarchical reinforcement learning, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.2020, pp. 1793-1796.
[30] H.-C. Kao, K.-F. Tang, E. Chang, Context-aware symptom checking for dis- ease diagnosis using hierarchical reinforcement learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, 20.