Technological Model using Machine Learning Tools to Support Decision Making in the Diagnosis and Treatment of Pediatric Leukemia

Daniel Mendoza-Vasquez, Stephany Salazar-Chavez and Willy Ugarte[a]
Universidad Peruana de Ciencias Aplicadas (UPC), Lima, Peru

Keywords: Model, Machine Learning, Leukemia, Decision Tree, Medical Assistance.

Abstract: In recent years, multiple applications of machine learning have been visualized to solve problems in different contexts, in which the health field stands out. That is why, based on what has been previously described, there is a wide interest in developing models based on machine learning for the creation of solutions that support medical assistance for disease such as pediatric cancer. Our work defines the proposal of a technological model based on machine learning which seeks to analyze the input medical data to obtain a predictive result, oriented to support the decision making of the specialist physician in relation to the diagnosis and treatment of pediatric leukemia. For the evaluation of the proposed model, a web validation system was developed that communicates with a service hosted on a cloud server which performs the predictive analysis of the inputs entered by the physician. As a result, an accuracy rate of 92.86% was obtained in the diagnosis of pediatric leukemia using the multiclass boosted decision tree classification algorithm.

1 INTRODUCTION

Pediatric cancer is one of the leading causes of death among children and adolescents worldwide. This disease affects everyone and can appear in any part of the body[1]. In addition, according to the World Health Organization (WHO) in its report made in 2018, it reports that in underdeveloped countries there is a lower survival rate of pediatric cancer due to the inability to obtain an accurate diagnosis, deaths from toxicity due to side effects caused by treatments and the lack of essential technological solutions[2].

Thus, it is detailed that, in Peru, 70% of pediatric cancer cases are diagnosed late, which hinders the development of a treatment. In addition, it is specifically detailed that the mortality rate of pediatric cancer is higher than 50%[3].

Currently, to diagnose and treat pediatric leukemia, specialists in pediatric oncology go through a complex process of integration of medical aspects to determine the recovery process, which includes radiotherapies, surgeries, physical examinations, among others. Consequently, there is an impact on the patient's health, since the procedure has a long duration (Marti-Bonmati et al., 2020). In addition, on the technological side, there are challenges such as the selection of the most appropriate classification algorithm to perform a predictive analysis and pattern identification to develop an alternative solution for the detailed problem.

Nowadays, there are projects that are developing models and solutions using technologies such as machine learning with cloud computing, applying different approaches, for example, the reading of medical images through computer vision for the detection of tumors, development of platforms that implement biomarkers to support diagnosis and treatment, among others (Marti-Bonmati et al., 2020; Fathi et al., 2020; Verda et al., 2019; Chaber et al., 2021).

Equally important, third level hospitals in Peru do not have the appropriate technology to optimize the time for diagnosis and treatment of pediatric leukemia, resulting in delayed evaluation and care [4]. As a consequence, the patient's recovery is complicated with little chance of recovery. With new trends and technologies, learning models make it possible to analyze this information using algorithms. Therefore, these models can be leveraged with information for analysis and prediction in pediatric leukemia.

[a] https://orcid.org/0000-0002-7510-618X

[1] PAHO "Childhood and Adolescence Cancer" - https://bit.ly/3jREaO6

[2] WHO "Childhood cancer" - https://bit.ly/3g0g1nA

[3] Gestion Newspaper - https://bit.ly/2Pa3weL

[4] Health Ministry of Peru - https://bit.ly/2XmHF7M

Mendoza-Vasquez, D., Salazar-Chavez, S. and Ugarte, W.
Technological Model using Machine Learning Tools to Support Decision Making in the Diagnosis and Treatment of Pediatric Leukemia.
DOI: 10.5220/0010684600003058
In *Proceedings of the 17th International Conference on Web Information Systems and Technologies (WEBIST 2021)*, pages 346-353
ISBN: 978-989-758-536-4; ISSN: 2184-3252

Machine learning is a process that uses automatic models that allow learning without any direct instruction. These learning models perform training that can last from minutes to hours (Rajagopal et al., 2020). Learning models can be adapted to different situations through classification and prediction. Thus, this technology has been used to classify and predict the different subtypes of leukemia. Thus, based on analysis and research, the boosted decision tree algorithm was selected to predict the diagnosis according to the type of pediatric leukemia. Consequently, this alternative solution will support the decision making of the pediatric oncology specialist, since, when making consultations, based on the history, signs, symptoms and laboratory results, in a system based on machine learning tools, a predictive result can be obtained to indicate whether the patient has the disease or not.

Our contributions are as follows:

- A tabular dataset has been constructed with input parameters such as medical history, signs, symptoms and laboratory results for diagnosis and treatment.
- A machine learning model has been designed and trained to be able to perform the classification according to the type of pediatric leukemia.
- A validation system has been developed with the aim of verifying how our proposal supports the pediatric oncology specialist in the diagnosis and treatment of pediatric leukemia.

This paper is has been organized as follows: Section 2 presents a brief theoretical framework of the most important concepts of our approach. Section 3 describes in detail our proposed technological model by explaining all the concepts involved in its development. Subsequently, in Section 4, we present the experiments carried out for the validation of our research and the results obtained. Section 5 gives details about the different existing research, and how in contrast with our proposal to generate a new solution. Finally, in Section 6, we present the conclusions and perspectives.

2 CONTEXT

In this section, the main requiered concepts for our approach will be introduced.

2.1 Pediatric Leukemia

This disease is among the most common cancers with the highest number of cases. Pediatric leukemia is a disease that originates in the bone marrow where new blood cells are formed and it has 4 types:

1. **Acute Lymphoblastic Leukemia (ALL)** is the most common cancer among children and the most frequent cause of death. This type of leukemia presents an abnormal increase of lymphoblasts in the patient (O'Brien et al., 2018).
2. **Chronic Lymphocytic Leukemia (CLL)** is a malignant neoplasm in which small, mature-appearing lymphocytes accumulate in the blood, bone marrow and lymphoid tissues.
3. **Acute Myeloid Leukemia (AML)** is a form of cancer characterized by the infiltration of bone marrow, blood and other tissues by proliferative hematopoietic systems cells (Naymagon et al., 2021).
4. **Chronic Myeloid Leukemia (CML)** represents 2% leukemias in children and adolescents. This disease presents with an increase in white blood cells and large spleen size (Hijiya and Suttorp, 2019).

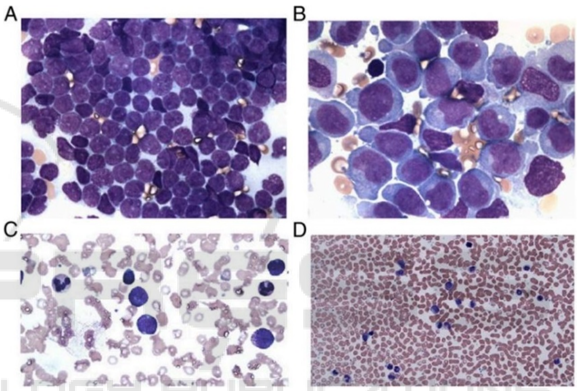

Figure 1: Microphotographs of the different type of leukemias (Kaplan, 2019).

Figure 1 shows the leukemic cells according to their type. In Figure 1A, ALL cells are evidenced where the large size of the nucleus to the cytoplasm is observed. In Figure 1B, CLL cells where the enlargement of the cytoplasm is visualized. In Figure 1C, AML cells are observed where the enlargement of the cytoplasm with prominent granules can be seen. In Figure 1D, CML cells where the number of myeloid progenitors in different stages can be observed.

2.2 Cloud Computing

Cloud computing is the compilation of integrated and networked hardware, software and internet infrastructure (Mathew et al., 2017).

1. **API services (application programming interfaces)** are interfaces that provide the program with the interaction with another system for data exchange.
2. **Azure functions** is a cloud service that provides infrastructure and resources for executings apps.

2.3 Machine Learning

This discipline belongs to artificial intelligence designed to create new systems that learn automatically without human intervention. Machine learning (ML) is a method of data analysis that automates the constructions of analytical models where this model can learn from experience to improve its performance (González et al., 2020).

A ML model has the following components:
1. **Dataset** is a collection of data that is made up of values from a table, which is part of database.
2. **A machine learning model** is a file that has been trained to recognize certain types of patterns.
3. **A Decision tree** is a classification technique is a tree structured representation where the nodes represent an attribute, the branches symbolize the test output and the final nodes are the classification result (Gonzalez et al., 2019).

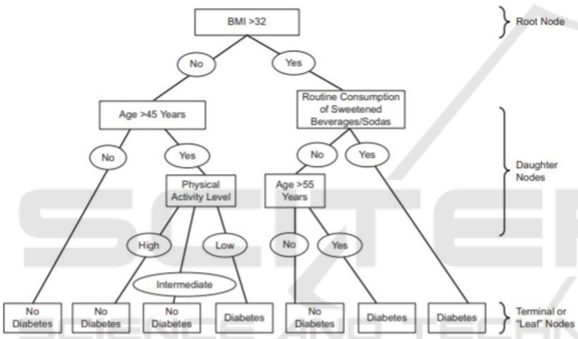

Figure 2: Decision tree classification model (Bi et al., 2019).

Figure 2 shows a case of a classification decision tree to predict a binary outcome for type 2 diabetes mellitus. Also, it should be noted that several works (Rodin et al., 2021; Zhao et al., 2021; Aslam et al., 2021) using this technique have used with different types of cancer providing a diagnostic result.

2.4 Validation

A the validation system has been built for the pediatric oncology specialist to provide support on the patient's diagnosis. In this way, input variables have been defined that allow the predictive analysis to obtain a positive or negative result of suffering the disease.

2.4.1 System Development

- **Framework:** It is a framework that enables software development. Therefore, Angular has been selected as the framework for this project.
- **Programming Language:** It is a set of instructions or actions designed for the execution of the system. For this purpose, TypeScript has been used since it is open source and based on JavaScript.
- **Execution Environment:** The execution environment used for the development of the project is Node.js, which is oriented to the interpretation of the JavaScript language.
- **Versioning:** The versioning and hosting of the system code has been done in the GitHub platform since it allows us to have a version control.

2.4.2 Input Parameters

For the development of the system, input parameters classified by medical history, signs, symptoms, and laboratory results have been defined to perform the predictive analysis with the objective of obtaining a result as a support suggestion in the diagnosis and treatment of pediatric leukemia.
- **Medical History:** The history is focused on the collection of patient information. Among the recorded histories of pediatric leukemia is DNA lesion, previous diseases, exposure to X-rays, chemotherapy, radiotherapy, among others.
- **Signs:** Signs are diagnosed during a physical examination with the patient. Among the signs that specialist identifies in pediatric leukemia are pallor, enlarged lymph nodes, fever, weight loss, appearance of red spots on the skin (petechiae), among others.
- **Symptoms:** The patient's symptoms are those ailments that the patient suffers from. The identified symptoms that the patient commonly suffers with the leukemia disease are extreme tiredness, night sweats, dizziness, blurred vision, among others.
- **Laboratory Results:** Laboratory tests analyze samples of blood or body tissue. With these results, the physician analyzes and determines the patient's current condition. For example, among the tests performed on the patient are the number of red and white blood cells, full blood count and others.

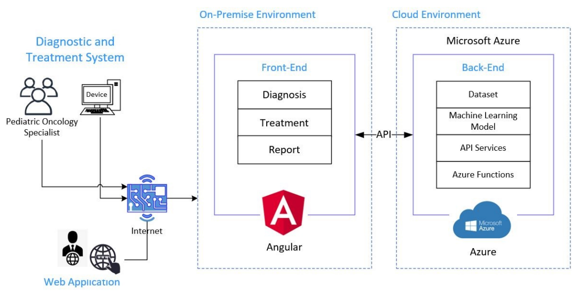

Figure 3: Integrated architecture of proposed solution.

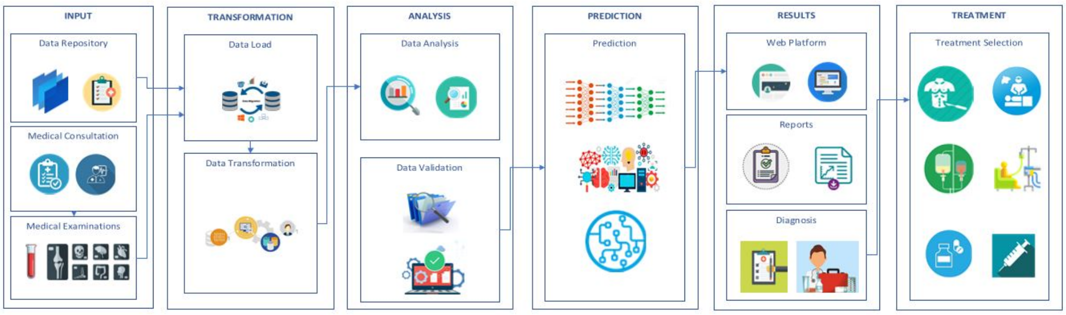

Figure 4: Proposed technological model for decision support in the diagnosis and treatment of pediatric leukemia.

3 TECHNOLOGICAL MODEL FOR PEDIATRIC LEUKEMIA

Based on the analysis of the current situation in Peru regarding the diagnosis and treatment of pediatric leukemia, the following solution proposal has been developed with the objective of providing support in decision making to the medical specialist, optimizing the time in the diagnosis and treatment of this disease. For this reason, the following paragraphs will detail the development of the proposal.

3.1 Integrated Architecture

The design of the integrated architecture that allows visualizing the integration of the development of the validation system to the technological model has been carried out. In this way, the pediatric oncology specialist has been identified as the main actor. This actor accesses the validation system with the objective of obtaining a diagnosis and treatment suggestion for pediatric leukemia (see Section 2.1).

Figure 3 defines the integrated architecture where the specialist enters through a device (computer) to the validation system via an Internet connection. The main actor visualizes the diagnosis and treatment module to obtain a predictive analysis report using the selected variables: history, signs, symptoms, and laboratory results (see Section 2.3) according to the type of leukemia (see Section 2.1) that the patient possibly suffers from. Also, there are two environments in our research proposal. The first environment is the local infrastructure where it is conformed with the front-end of the validation system. Thus, this system has the diagnosis and treatment and reporting modules. It should be noted that in the reports the percentage of accuracy obtained with the machine learning model is displayed. The second environment is the cloud infrastructure, where it is conformed by the Azure services used for the development of the model.

It should be noted that the information is kept secure through services offered by Azure platform such as data encryption, access management and implementation of audits to ensure compliance with the security policies established by the client.

3.2 Technological Model

The following technological model (see Figure 4) has been developed to improve the accuracy in the diagnosis and treatment of pediatric leukemia. The objective of this model is to optimize and improve medical care for the patient. For this reason, the following modules have been determined.

1. **Input:** This phase consists of three modules:
- The data repository is focused on the patient's medical data and important documents.
- The medical consultation is focused on the medical evaluation of the patient.
- The medical examinations that the patient must undergo to determine his or her diagnosis.

2. **Transformation:** This phase is focused on data transformation. Therefore, it consists of two phases: The first module is the data loading where the preparation of the information in the previous phase is performed. The second module is the data transformation, which adapts the data conversion for its analysis.

3. **Analysis:** This phase is oriented to the analysis of data collected in the previous phase. Therefore, two modules have been defined in this stage: The data analysis module is focused on the verification of these obtained from the patient. The second module is the validation of the data, where the correctness of the data is confirmed.

4. **Prediction:** This phase is focused on performing pediatric leukemia diagnosis and treatment prediction using Azure Machine Learning tool using supervised classification algorithms.

5. **Results:** This phase is focused on visualizing the reports with the results obtained through the predic-

tive analysis model. Thus, it has been segmented into the following modules. First, there is the web platform, which is the medium where the specialist interacts to identify and visualize the results obtained. Secondly, it is focused on the reports where the accuracy obtained for the diagnosis of pediatric leukemia is detailed. Finally, there is the definitive diagnosis where it is indicated whether the patient may suffer from this disease based on the predictive analysis performed.

6. **Treatment.** This last phase consists of providing treatment suggestions based on the results of the diagnosis provided by the predictive analysis tool. Thus, the specialist analyzes the suggestions and determines the ideal treatment for the patient's recovery.

3.3 Development of the Predictive Analytics Model

Based on the design of the proposed technological model, its development was carried out at a functional level through the Azure Machine Learning cloud platform, which offers tools as a service could allow obtaining a predictive result in relation to the diagnosis of pediatric leukemia. Thus, the following lines will detail the process of creating the predictive model:

Firstly, a resource group was created, which refers to a container of services provided by the Azure Cloud Platform, which contains virtual networks, virtual machines, coding services, machine learning and other. In this way, we started by creating a dataset in a spreadsheet, with the values of each input variable. Also, it is highlighted that the input medical parameters represented by general analysis, signs, symptoms, and laboratory results were extracted from research and reports of formal medical institutions.

Secondly, after having the dataset, we proceeded to the creation of the environment where the predictive analysis is going to be performed, for which reason we used the service called process, for the generation of a resource represented by a virtual machine which will allow the training of the model.

Thirdly, we proceed to the design of the predictive model, where through the components provided by Azure such as machine learning algorithms, data analysis, statistical functions, training model and others, it is possible to establish the flow that will allow to obtain a result. Therefore, the first component is the parameterized dataset, then there is the data normalization component to allow the algorithm to model the data correctly. Then, the training and test data were separated (with the split data component). After, the MultiClass Boosted Decision Tree algorithm was selected to perform the prediction, since there are four types of leukemia to be classified. Next, the training phase is performed, which allows learning from patterns identified based on historical data sets (with the train model component). Finally, a score and an evaluation of the training (with the components score model and evaluate model) were obtained, which will be detailed in Section 4.

3.4 Validation System

To verify the performance of the model, a web system was developed under a controlled environment using the dataset with test cases established by the researchers detailed in the previous section. As part of our validation system, the pediatric oncology specialist selects the patient along with the type of leukemia he/she may possibly have. In this way, the various primers are displayed with the variables that the patient may present. As a result, the diagnostic report is displayed indicating the patient's status (positive or negative) with the percentage of accuracy or false positives based on the analysis of the predictive model developed in Microsoft Azure Machine Learning. Also, the validation system displays the "Treatment" option when the result is positive.

4 EXPERIMENTS

In this section, the experiments carried out in this paper will be explained, starting from the experimental protocol and the results, as well as its discussion.

4.1 Experimental Protocol

For the development of our research project, several activities have been carried out to obtain an expected result. Figure 5 details the steps that have been carried out for the implementation of our approach.

4.1.1 Dataset Creation

A dataset has been created to obtain the percentage of accuracy in the diagnosis of the four leukemia subtypes (see Section 2.1). Then, we have placed the input values and generated the casuistry to train the model and thus obtain a predictive result with a percentage of accuracy of having or not having the pediatric leukemia disease. It should also be noted that the input parameters were collected from medical research and validated by a pediatric specialist.

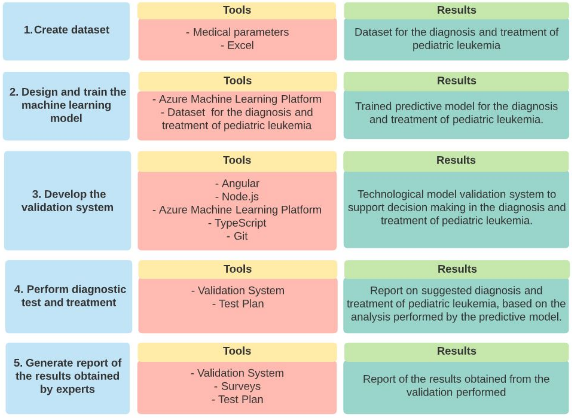

Figure 5: Proposed technological model for decision support in the diagnosis and treatment of pediatric leukemia.

4.1.2 Design and Training of the Predictive Model

A model has been developed using the tools of the Azure Machine Learning cloud platform, from which training has been performed with the input values with certain controlled casuistries to obtain the percentage of accuracy. Therefore, to train the model, the parameterized dataset with the medical values developed in the previous phase was used. It should also be noted that the following Azure services have been used to train the model:
1. **Process:** This service allows to use the resource as a destination to perform the training of the model with the following specifications:
- Process name: ML1-Process
- Process type: Machine Learning compute
- Operating system: Linux 3.10 x86_64
- Processor: Intel Xeon E5-2673 @2.29GHz 4 cores
2. **Pipelines:** This Azure Service allows to organize step by step the machine learning flow in parallel with the data processing with the following results:
- Training duration: 1 hour, 8 minutes.
- Process destination: ML1-Process.
- Accuracy: 92.86%
3. **Validation System:** This validation system has been elaborated under the approach of a controlled environment with the processing of simulated data to obtain a predictive result provided by the Azure tool with the following technologies:
- Framework: Angular
- Programming language: TypeScript
- Execution Environment: Node.js
- GitHub: https://bit.ly/2UhUqzm

In addition, to collect and process the data, an API was developed using Azure Functions to enable communication between the web app and the model.

4.2 Results

4.2.1 Treatment and Caring Time

In addition, in Lima, the delay from the first symptoms of cancer to access to a hospital specialized in pediatric oncology care was 81 days. Likewise, in the regions of origin, the delay time from the appearance of the first symptoms was 63 days (142 days in ESSALUD). On the other hand, in Lima, the delay time is 61 days (88 days in MINSA)[5]. It is evident that currently, the specialist has a delay time in the diagnosis of pediatric leukemia of 1464 hours (61 days). Therefore, based on the test performed together with experts, it can be pointed out that, the specialist will take a maximum time of 10 additional minutes to obtain an accurate suggestion to support decision making in the diagnosis and treatment of pediatric leukemia.

4.2.2 Validation with Experts

For our work, validation has been carried out with experts in the field of pediatric medicine and development of health care systems. In this way, meetings were held with experts where the research proposal was detailed. For further details, please see the following link https://bit.ly/3zikrhq.

Likewise, the validation system has been presented to the experts indicating its functionalities to demonstrate the predictive result in support of the diagnosis and treatment of pediatric leukemia. As a result, the specialist has performed the corresponding tests on the validation system in a time interval of 8 to 10 minutes based on the developed test plan that can be visualized in the following link https://bit.ly/3pFGiuI (in Spanish).

4.2.3 Algorithm Results

As part of the experiment of this research, different supervision algorithms have been used to compare the results obtained in relation to the percentage of accuracy. Table 1 shows the percentage obtained according to the type of classification algorithm.

Table 1: Accuracy score w.r.t the classification technique.

Classification algorithm	Percentage of accuracy
Multiclass Decision Forest	23.61%
Multiclass Logistic Regression	33.33%
Multiclass Neural Network	85.71%
Multiclass Boosted Decision Tree	92.86%

[5] Peruvian ombudsman - https://bit.ly/363qPf5

Thus, based on the experiment carried out, it was validated that the result obtained by the algorithm called “Multiclass Boosted Decision Tree” is the most suitable for predictive analysis based on the input medical parameters. Therefore, this algorithm was selected to provide a suggestion result with a high degree of accuracy, which supports the diagnosis and treatment of pediatric leukemia.

4.2.4 Diagnostic Error Rate

On the other hand, the misdiagnosis rate has identified opportunities to reduce the recovery rate of pediatric cancer. Therefore, the key factor of the research factor of the project is to focus on the lack of diagnosis or misdiagnosis since approximately 20% of children are affected with this diagnosis[6]. Table 2 shows the percentage of error represented by false positives according to the classification algorithm provided by Microsoft Azure Machine Learning. As a result, it is observed that the neural network algorithms and the boosted decision tree algorithm have a lower percentage of error compared to the percentage of error in the diagnosis of pediatric cancer.

Table 2: % of false positives w.r.t. classification technique.

Classification Algorithm	Percentage of accuracy
Multiclass Decision Forest	76.39%
Multiclass Logistic Regression	66.67%
Multiclass Neural Network	14.29%
Multiclass Boosted Decision Tree	7.14%

Thus, it is demonstrated that with the neural network and boosted decision tree algorithms there is a lower percentage of false positives compared to the diagnostic error rate of 20% detailed in the PAHO report. Thus, based on the above, we have a viable solution to support specialists in making a better decision in the definitive diagnosis of pediatric leukemia.

4.3 Discussion

At the end of the result stage, the following discussions have become evident in relation to the development of this research. Based on the development of the technological model validation system, it was possible to verify that the process of obtaining a predictive result in relation to the diagnosis and treatment of pediatric leukemia has a time interval of between 8 and 10 minutes. In addition, to have the ability to determine the most appropriate pediatric oncologic treatment line for the patient's recovery. The ideal classification algorithm for the development of this project is the boosted decision tree algorithm base on the comparison of the results of other algorithms (see Table 1), since an accuracy percentage of 92.86% was obtained. It should be noted that this percentage was obtained based on the use of the Microsoft Azure Machine Learning platform, so this research seeks those other studies use different platforms and design multiple machine learning models to obtain better results. The present research proposal aims to support the process of diagnosis and treatment of pediatric leukemia by reducing the percentage of misdiagnosis to increase the recovery rate of the disease. For this reason, the results of false positives have been analyzed and detailed to solve the problem of misdiagnosis (see Table 2).

5 RELATED WORKS

In (Fathi et al., 2020), the authors propose an expert system based on neural networks with the aim of performing the prognosis and classification of the type of leukemia in children according to complete blood count test, ANFIS (Artificial Neural Network Fuzzy Inference System), GMDH and metaheuristic algorithms. Therefore, this proposal is conformed in the collection of data and samples, training and verification of answers, data classification and data division. As a part of the study results, it has been demonstrated that there are limitations in the separation of cancer types due to the high percentage of error (Fathi et al., 2020). In relation to our proposal, our model is focused on the patient's symptomatology as input values to make the diagnosis of pediatric leukemia according to its type allowing to visualize treatment suggestions for the early recovery of the patient.

In (Verda et al., 2019), the authors detail the development of Logic Learning Machine (LLM) in order to perform gene expression data analysis for pediatric cancer diagnosis. In this research, a comparison has been made with existing supervised analysis methods such as decision tree (DT), artificial neural network (ANN) and k-means classifier algorithms. For this purpose, the authors have used a dataset of eight databases including cancer cell types. In contrast to our research proposal, our project is focused on the four subtypes of pediatric leukemia using the multiclass classification algorithm. To make the diagnosis of this disease, we have analyzed the patient's history, signs, symptoms, and laboratory results without focusing on the patient's genetics. In (Chaber et al., 2021), the authors mention the use of the Fourier

[6]Health Ministry of Peru “Early Diagnostic Guide Cancer in Children and Teens in Peru” - http://bvs.minsa.gob.pe/local/MINSA/5044.pdf

transform infrared spectroscopy (FTIR) tool as it is a fast and cost-effective method, which allows early detection of cancer-specific chemical changes in tissues, cells and biofluids. In this research, it is proposed to use this tool for diagnosis in pediatric acute lymphoblastic leukemia using blood samples. As part of the experiment, an evaluation has been performed with 10 patients with this disease. Given this, the authors developed a predictive model based on AdaBoost with a percentage of 85% accuracy. In contrast to our approach, a technological model has been developed focused on the diagnosis and treatment of the four subtypes of pediatric leukemia through laboratory results, symptoms, signs, and general medical aspects of the patient. From which, an accuracy of 92.86% was obtained.

6 CONCLUSIONS

A machine learning model trained with a dataset in a tabular manner with medical history, symptoms, signs, and laboratory results has been developed to be able to identify whether the patient has high probability in suffering from pediatric leukemia disease.

It has been shown that the multiclass boosted decision tree algorithm has a high percentage of accuracy (92.86%) for obtaining a predictive result of suggestion to support the diagnosis and treatment of pediatric leukemia. In addition, there is evidence of an opportunity to reduce the misdiagnosis results from the solution, since a lower percentage of false positives (7.14%) was obtained.

An interesting future work can be the analysis of information such as medical history (with machine learning), medical images (with computer vision) or adding new modules to complement and increase knowledge and support the recovery rate of pediatric cancer disease or its protection with blockchain (Arroyo-Mariños et al., 2021).

REFERENCES

Arroyo-Mariños, J. C., Mejia-Valle, K. M., and Ugarte, W. (2021). Technological model for the protection of genetic information using blockchain technology in the private health sector. In *ICT4AWE*.

Aslam, M., Xue, C., Chen, Y., Zhang, A., Liu, M., Wang, K., and Cui, D. (2021). Breath analysis based early gastric cancer classification from deep stacked sparse autoencoder neural network. *Scientific Reports*, 11.

Bi, Q., Goodman, K., Kaminsky, J., and Lessler, J. (2019). What is machine learning: a primer for the epidemiologist. *American journal of epidemiology*, 188.

Chaber, R., Kowal, A., Jakubczyk, P., Arthur, C., Łach, K., Wojnarowska-Nowak, R., Kusz, K., Zawlik, I., Paszek, S., and Cebulski, J. (2021). A preliminary study of ftir spectroscopy as a potential non-invasive screening tool for pediatric precursor b lymphoblastic leukemia. *Molecules*, 26.

Fathi, E., Jahangoshai Rezaee, M., Tavakkoli-Moghaddam, R., Alizadeh, A., and Montazer, A. (2020). Design of an integrated model for diagnosis and classification of pediatric acute leukemia using machine learning. *Journal of Engineering in Medicine*, 234.

González, E. E. M., Reyes-Ortíz, J. A., and González-Beltrán, B. A. (2020). Machine learning models for cancer type classification with unstructured data. *Computación y Sistemas*, 24(2).

Gonzalez, G., Lu, Z., Leaman, R., Weissenbacher, D., Boland, M. R., Chen, Y., Du, J., Fluck, J., Greene, C. S., Holmes, J. H., Kashyap, A., Nielsen, R. L., Ouyang, Z., Schaaf, S., Taroni, J. N., Tao, C., Zhang, Y., and Liu, H. (2019). Text mining and machine learning for precision medicine. In *PSB*.

Hijiya, N. and Suttorp, M. (2019). How i treat chronic myeloid leukemia in children and adolescents. *Blood*, 133.

Kaplan, J. (2019). Leukemia in children. *Pediatrics in Review*, 40.

Marti-Bonmati, L., Alberich-Bayarri, A., Ladenstein, R., Blanquer, I., Segrelles, D., Cerdá-Alberich, L., Gkontra, P., Hero, B., Garcia Aznar, J. M., Keim, D., Jentner, W., Seymour, K., Jiménez-Pastor, A., González Valverde, I., Heras, B., Essiaf, S., Walker, D., Rochette, M., Bubak, M., and Neri, E. (2020). Primage project: predictive in silico multiscale analytics to support childhood cancer personalised evaluation empowered by imaging biomarkers. *European Radiology Experimental*, 4.

Mathew, S., Gulia, S., Singh, V., and Dev, V. (2017). A review paper on cloud computing and its security concerns. In *RICE*, volume 10.

Naymagon, L., Najfeld, V., Teruya-Feldstein, J., El Jamal, S., and Mascarenhas, J. (2021). *Acute Myeloid Leukemia*.

O'Brien, M., Seif, A., and Hunger, S. (2018). Acute lymphoblastic leukemia in children. *Wintrobe's Clinical Hematology: Fourteenth Edition*.

Rajagopal, S., S., H., and Kundapur, P. (2020). Performance analysis of binary and multiclass models using azure machine learning. *International Journal of Electrical and Computer Engineering (IJECE)*, 10.

Rodin, A., Gogoshin, G., Hilliard, S., Wang, L., Egelston, C., Rockne, R., Chao, J., and Lee, P. (2021). Dissecting response to cancer immunotherapy by applying bayesian network analysis to flow cytometry data. *International Journal of Molecular Sciences*, 22.

Verda, D., Parodi, S., Ferrari, E., and Muselli, M. (2019). Analyzing gene expression data for pediatric and adult cancer diagnosis using logic learning machine and standard supervised methods. *BMC Bioinform.*, 20(9).

Zhao, Y., Chang, C., Hannum, M., Lee, J., and Shen, R. (2021). Bayesian network-driven clustering analysis with feature selection for high-dimensional multimodal molecular data. *Scientific Reports*, 11.