*Corresponding Author's Email: flabasouza@yahoo.com.br
Proceedings of the International Academic Conference on Teaching, Learning and Education
Vol. 3, Issue. 1, 2026, pp. 28-42
DOI: https://doi.org/10.33422/tleconf.v3i1.1869

ISSN: 3030-0614 online

The Use of Artificial Intelligence in Medical Education: A Case Study

Flavio Henrique Batista de Souza*, Bráulio Roberto Gonçalves Marinho Couto, and Ana Paula Gandra Martins
Centro Universitário de Belo Horizonte UNIBH. Brazil

Abstract

Artificial intelligence (AI) has been increasingly used by the population and also in academia, and can promote different impacts in the field. AI is a mechanism that aims to match the human learning capacity, being able to predict responses through learning algorithms: Machine Learning. Just as humans are proposing to connect more with these tools, medical science is also moving in the same direction. In this regard, it is relevant to discuss alternatives that consider and experiment with such use and avoid, above all, errors. Thus, this research aims to present a case study of the use of AI algorithms in the area of medical education, report projects already carried out and present impacts of the proposed tool. This case study focuses on teaching Medical Informatics to 180 students in the first period of the medical course using hospital databases and algorithms. As results, 20 publications were obtained in international conferences, more than 180 students, freshmen with impact publications (in A1 qualis journals, or impact factor greater than 2). With experience in the practice of data analysis and management using innovative tools within AI medical education, in addition to an experience of differentiated scientific production by students in the first semester. Despite the success of the experiments, such as AUCs of 0.98 and publications, relevant issues such as hard skills assessments have become pertinent in order to evaluate educational gains. The use of AI as a learning and scientific production mechanism within medicine offers different academic and professional benefits. Therefore, more projects like those reported should be encouraged in the medical field.

Keywords: Medical Informatics; Medical Education; Artificial Intelligence; Machine Learning

1 Introduction

The use of artificial intelligence (AI) in education has been widely discussed by various academic perspectives. Not only in its form of application, but also in what results could be obtained, in addition to the real impact on student training (Cruz et al., 2023). The question of whether it behaves as a contribution or a pejorative issue in its use during the student's journey has been highlighted in academic circles.

In the first instance, AI can be considered an area of study where the premise of decision-making capacity is emulated by an artificial means, in this case, computers. For this purpose, it relies on machine learning algorithms, that is, Machine Learning algorithms, where mathematical structures are used to emulate the learning process of a human being (Cruz et al., 2023; Souza & Morão, 2020). Among these learnings, one of them, which has a high connection with management processes and medical decision-making, is the prediction process, where the algorithm aims, according to a certain accuracy, to predict a response to a problem (or condition), in view of what the algorithm "learned" from past data (Souza & Morão, 2020).

Given a scenario where we have human capacity proposing to connect to computerized systems, we have medical education. Always considered traditional, medical education brings with it methodologies with decades of references and experiments. However, over the last few years, it is noticeable how medical science, not only in its essence but also in the university environment, has been grappling with elements of computer technology, information systems, computerized equipment, and examples of the presence of medical technology. This shift is highlighted by the critical need for Data Literacy competencies in medical curricula, which is increasingly recognized as essential for future clinical decision-making and patient management (Hoffmann et al., 2022). Therefore, implementing hands-on training using real-world data and algorithms, such as Machine Learning, is crucial for developing these essential analytical skills in undergraduate medical students (Lee et al., 2021).

On this aspect, we have the case of the application of AI in studies, where its use must be carefully considered so that, as described by Ricieri et al.(2024), errors (even their recurrence) are avoided. Beyond the technical challenges of AI implementation, there is a fundamental pedagogical gap regarding robust assessment frameworks for educational innovation. The literature, particularly within the Scholarship of Teaching and Learning (SoTL), emphasizes that changes in teaching methodologies must be systematically investigated and proven through measurable learning outcomes, not just anecdotal evidence (Hutchings & Shulman, 1999; Gansemer-Topf et al. 2024). Therefore, our approach is not only to report the use of AI, but also to validate its educational impact through rigorous pre/post measures, directly addressing the demand for validated hard skills assessments in technology-integrated medical education.

Thus, the general objective of this article is to present a case study, coordinated by two technology professors who work on medical education, on the use of AI algorithms as a way to connect content and analyze data. The specific objectives are described in two categories:

- **Description of the experimental processes of real clinical data**: focus in demonstrate the process of connection between the teaching of Medical Informatics and AI tools; report some of the projects already demonstrated; present a compilation of publications.
- **Measurable Learning Objectives:** assess acquired knowledge, where the student should be able to describe, with 80% accuracy, the functions and types of categorical and quantitative variables in a hospital data dictionary; verify technical skill, where the student must demonstrate technical proficiency (competence) in manipulating the RStudio® software to apply the Multi Layer Perceptron (MLP) algorithm in the prediction of Surgical Site Infection (SSI); demonstrate critical reasoning, where the student will justify, in their final report, in the format of a scientific article, the decision to exclude a minimum of 10% and a maximum of 60% of the noisy records from their database, using descriptive statistics (mean, median, maximum and minimum) as a basis.

This research is justified by the high demand for scientifically based work on the impacts (positive and negative) of using the proposed tool.

2 Methods

The proposal was carried out as follows: at the beginning of the semester, in the Medical Informatics course, within the medical program, the technologies currently used in the medical field were presented, and the proposal was made that the final project would be an experiment, using real hospital databases collected between 2016 and 2018, to test the use of machine learning algorithms for predicting surgical site infection. An algorithm was selected, and several databases relating to various types of surgeries were subjected to the experiments.

This project began in 2018 with first-year medical students. The databases used are data collected between 2016 and 2018 in various hospitals. They include the anonymization process of patients, focusing only on data related to care and the surgical process, such as surgery duration, sepsis, age, weight, among others. Several surgeries were tested, totaling 18, such as craniotomy, orthopedic surgeries, vascular surgery, among others. Focusing on the teaching methodology, the study project based on Machine Learning algorithms, which support AI-based applications, followed the following application sequence:

1- Data acquisition: Each semester, a set of specific surgery databases, with real data from hospitals, which were duly submitted to all the rigors of the professional Ethics Committees, both from the university and the hospitals, were selected so that medical students could carry out analyses and experiments. Regarding the data acquisition processes and the relevant ethics committees, it should be noted that the data were validated through research projects approved on the Plataforma Brasil platform (https://plataformabrasil.saude.gov.br/login.jsf), which can be identified by the following CAAE numbers: 39428114.9.1001.5093; 39445914.8.0000.5093; 60652416.1.1001.5093; 60652216.7.1001.5093.

2- Team formation: the teams were already formed, focusing on groups of 6, at most 7 medical students to receive and work on a specific database. There were no preferences regarding the students elected by group. At this time, there was an explanation of how a data collection process works, professional ethics committees, and the responsibility of what was being proposed.

3- Experimentation period: the students had to carry out an analysis and experimentation process, bringing a report in the form of a scientific article, describing all the processes that were carried out and the results obtained.

4- Pre-alignment: Each group had a contact student to receive the experimental data. Their reference data for correspondence were those already provided by the university. The anonymized databases were manipulated exclusively using R® software in a controlled laboratory environment. Access to the Excel® files was limited to class time and monitored by the professors.

5- Meetings: Guidance sessions were scheduled, with 30 minutes per week of each class dedicated to each group of students for guidance. Attendance was recorded in each class, and the results of each meeting were evaluated, with partial submissions of each group's articles, generating an average of 5 submissions per group per semester, where scores were awarded from 0 to 10 points per submission.

6- Analyses and experiments: The analyses involve an evaluation of the database, each containing a data dictionary, as exemplified in Table 1. Each database contains 27 variables, including quantitative and descriptive ones. The databases, as described in the published articles, could contain more than 40,000 records with each of the 27 variables. Values such as mean, median, maximum and minimum values were the references for data cleaning decision-making. The evaluation through algorithm experiments followed the parameter variations described previously. The balancing of the databases was not a constant, since the focus is on a real database, with all its costs and problems. It should be noted that the database, during its cleaning and adaptation process, may contain data from processed variables. For example, the Medical Record was excluded during the analysis process as a form of anonymization.

Table 1: Examples of variables and descriptions of databases (Variables 1 to 7 from 27)

Variable	Description	Function	Variable type
Medical Record	Patient's medical record number	Identification variable	Not applicable
Type_of_Surgery	Type of surgery performed	Auxiliary variable	Not applicable
Hospital	Hospital code where the surgery was performed (A, B, C, ...)	Explanatory variable	Categorical
Hospitalization_Number	Number of patient hospitalizations (1, 2, 3 ...)	Explanatory variable	Quantitative
First_Hospitalization	Was the surgery performed during the patient's first hospitalization? (Yes or No)	Explanatory variable	Categorical
Age_Years	Patient's age (years)	Explanatory variable	Quantitative
Above 70 Years General	Patient over 70 years of age? (Yes or No)	Explanatory variable	Categorical

Source: (Authors)

For data quality control (Data Cleaning), exclusions Rules were established:

- **Records:** A record (patient) was entirely removed if more than 5 of the 27 variables were missing or if a critical quantitative variable (e.g., "Age_in_Years" or "Surgery Duration") was outside ± 3 standard deviations of the group mean.
- **Variables:** If a reference variable had 40 to 80% noisy data (incorrect entries, empty spaces, unrealistic values), an entire column of that variable would be removed.

7- Pre/Post Measures - assessment of familiarity with the resources of the experiments: Before starting the experiments, a pre-assessment with the students would be carried out:

- *Are you proficient with computers? (yes or no)*
- *Do you use or have you used Microsoft Excel or Google Sheets? (yes or no)*
- *Do you have any basic knowledge of programming? (yes or no)*
- *Are you comfortable with work involving computers? (yes or no)*
- *Do you understand that data analysis is relevant to your profession? (yes or no)*
- *Are you familiar with scientific articles? (yes or no)*

This pre-questionnaire would be followed by the post-assessment:

- *Did you effectively participate in the computational part? (yes or no)*
- *Did you understand the meaning of the data in your database? (yes or no)*
- *Did you understand the results of the MLP algorithm? (yes or no)*
- *Have you ever attended or are you familiar with scientific conferences? (Yes or No)*

- *Technical Skill (RStudio® and Data Cleaning): report of which variables were removed and how many records remained, with consistency exceeding 80% proximity to a clean dataset.*
- *Critical Reasoning (Justification of Accuracy vs. Imbalance): present at least one coherent justification of the accuracy (AUC) versus the imbalance of the dataset.*

8- Evaluation and Grades in the semester: Student submissions were evaluated from 0 to 10, at each of the 5 meetings proposed per semester. Each section of the article would correspond to the evaluation report, being: Abstract, Introduction, Materials and Methods, Results, Conclusion (References were evaluated at all meetings). Table 2 describes the rubric used to evaluate each group, according to the delivery of each section of the article.

Table 2: Evaluation Rubric

Section	Measurements	Evaluation (0-10)
Introduction	Does it include Context, Problem Statement, General Objectives, Specific Objectives, and Justification? Are citations properly provided?	0-10
Methods	Are reference frameworks, experimental techniques, and parameters present? Are the analyses and descriptions adequately described?	0-10
Results	Does the section include a description of the results of the Experiments, Weighted Analyses, and Discussions about the results presented?	0-10
Conclusion/References	Are the Assessment, Findings in Relation to Objectives, and APA standards present and aligned?	0-10
Attendance at meetings	Attendance at all meetings?	0-10

**Note: The detailed descriptive evaluation criteria for each item in the Rubric (0-10) are internal course documents, and this table focuses only on the evaluated elements of the final paper.*

Source: (Authors)

9- Publication and co-authorship process: the professors involved selected the international conferences that had the best fit for the research carried out. All students in the group were considered co-authors. During the submission process, all students and the project's coordinating professors who participated in the data collection process (who were not students and submitted the data) were listed as authors of each article. The authorship criteria references the International Committee of Medical Journal Editors (ICMJE)

The evaluation process with pattern recognition algorithms, for predicting surgical site infection, the Multi Layer Perceptron algorithm was used. This algorithm is described as follows: it is based on learning by error correction. According to Souza & Mourão(2020):

> "*The Backpropagation algorithm can be classified as a type of supervised learning that calculates the error generated in the output values and propagates them to the input in order to adjust the weights and calculate the values again. In an MLP network architecture, a vector is presented to the nodes of the input units x of the network and its effect propagates through it, passing through all layers until it reaches the output units y, which is produced as the network's response.*"

Figure 1 shows an example of the MLP structure.

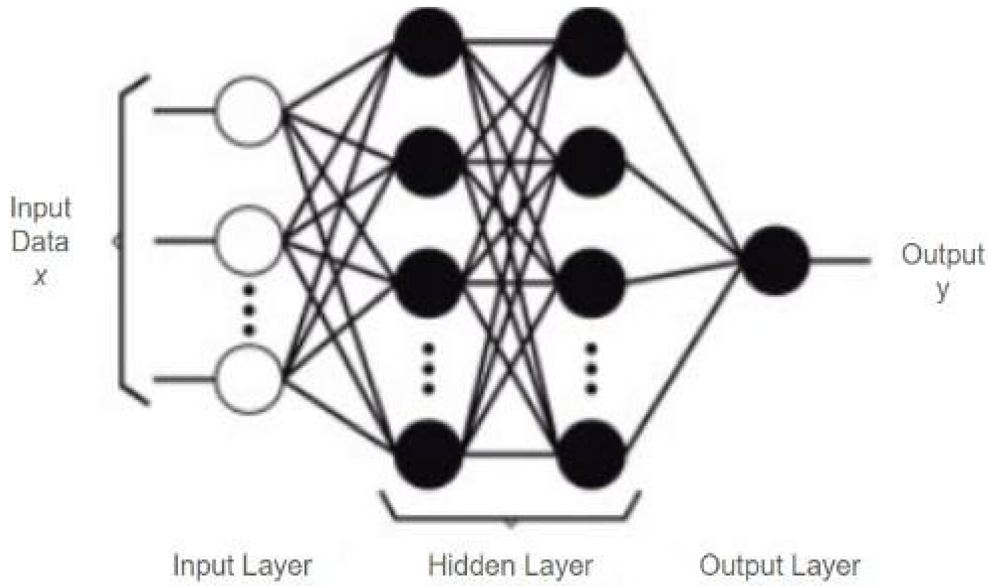

Figure 1: Example of MLP structure
Source: (Souza & Mourão,2020)

During the experimentation process, parameters such as the number of neurons in the hidden layer (3, 5, and 7), the percentage of the database for training and validation (65/35% and 75/25%), and variations in the type of MLP structure (Quick Propagation, Backpropagation Standard, momentum, resilient propagation, and Weight Decay) were varied. The values for the number of neurons (3, 5, and 7) and the resampling percentage were limited to those configurations that presented acceptable convergence time for the RStudio® in the laboratory environment, ensuring that students could complete the full cycle of 150 experiments within the established class period (30 minutes per group, weekly).

To measure the accuracy of the MLP evaluation on each of the databases, the Area Under the Curve (AUC) measure was used. A measure between 0 and 1 of an area generated by the ROC (Receiver Operating Characteristic Curve), where the measure of the area under the curve generates the AUC. Figure 2 demonstrates an example of a ROC curve drawing that should be taken as referece in the AUC calculation process. The closer to a perfect square shape, the higher the AUC, the closer to 1, and the better the algorithm's accuracy.

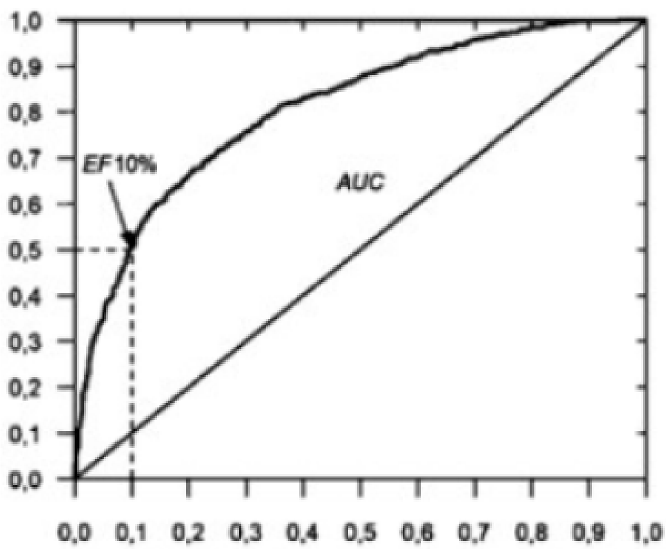

Figure 2: Example of ROC Curve and AUC
Source: (Souza & Mourão, 2020)

The imbalance in the databases was significant, with the incidence of SSI ranging from 5% to 15% among the different types of surgery. AUC was chosen as the primary performance metric because it is robust to imbalanced databases, avoiding sensitive metrics such as simple accuracy.

The evaluation process of the MLP model was based on the simple random sampling technique with fixed split (Holdout Method), instead of cross-validation, due to the processing time constraints inherent in the didactic laboratory environment and the need for students to complete 150 experimental combinations within the established class period. To ensure robustness and mitigate the potential bias introduced by simple splitting, the following procedures were adopted:

- **Randomness and Shuffling:** Before each split, the complete (already cleaned) dataset was fully shuffled in a pseudo-random manner.
- **Stratified Partitioning (Implicit in Randomization)**: The database was divided into two proportions: 65% for training and 35% for validation, and 75% for training and 25% for validation. Prior shuffling ensures that the proportion of the target variable (SSI – Surgical Site Infection), which ranged from 5% to 15% in the datasets, was consistently preserved in both subsets.
- **Multiple Test Rounds:** Each experiment, with its specific combinations of hyperparameters (neurons, MLP type), was run in five independent rounds. This allowed for obtaining the average Area Under the Curve (AUC), a more stable and representative metric of performance, minimizing the influence of a single split and attesting to the consistency of the results found by the students.

The students used the RStudio® software to perform their experiments during their laboratory classes. The project was carried out with over 180 medical students in their first semester, during the semesters of 2018/2, 2019/1 and 2019/2.

The implemented medical informatics teaching methodology, which requires first-semester students to complete a comprehensive data analysis project using Machine Learning (ML) algorithms, must be framed within robust pedagogical structures. Instead of a traditional didactic approach, the project adopts a Project-Based Learning (PBL) format, where students engage in solving complex real-world problems – in this case, predicting surgical site infection (SSI) in hospitals (Thomas, 2000; Kokotsaki, et al. 2016). This focus on knowledge creation, rather than just consumption, aligns directly with the principles of the Scholarship of Teaching and Learning (SoTL), which advocates that pedagogical excellence be investigated and proven through measurable learning outcomes (Hutchings & Shulman, 1999; Gansemer-Topf et al. 2024).

3 Results

3.1 Publications Achieved through the Results of the Experiments

During the semesters of 2018/2, 2019/1 and 2019/2, first-semester medical students, in the Medical Informatics course, are introduced to the concepts of databases, types and concepts of variables and statistics. After a theoretical foundation on these subjects, the advances in technology through mathematical algorithms are mentioned.

After understanding these structures, students are guided through a practical experimentation process, where databases, in Excel® format, are distributed among groups of 6 to 7 students. These databases contain data collected from 2016 to 2018, focused on surgical site infection

(SSI). These databases, which consolidate information from 3 to 7 hospitals, cover the following surgeries as shown in Table 3:

Table 3: Types of Surgeries from each Database

Groups of 2018/2	Groups of 2019/1	Groups of 2019/2
General	Knee Arthroplasty	Aortic Aneurysm
Cardiac	Hip Arthroplasty	Colon Surgery
Plastic Surgery	Limb Amputation	Prosthetic Surgery
Orthopedic Surgery	Bariatric Surgery	Emergency Surgery
Craniotomy	Cholecystectomy	Infected Surgeries
Vascular	Pacemaker Implantation	Clean Surgeries
Obstetrics	Abdominal Hysterectomy	Small Intestine
-	Revascularization	Exploratory Laparotomy
-	-	Patients Over 70 Years Old
-	-	Orthopedic Trauma

Source: (Authors)

Each database has an average of 27 variables, such as weight, height, surgery time, among others. In addition, a specific variable is the SSI (whether it happened or not). Before submitting the database to the MLP evaluation process, the students performed the database cleaning process. They needed to make decisions such as: removing the entire record from a collection (1 patient and all their 27 pieces of information according to the variable) or an entire variable (with its thousands of records), where there were databases with an original collection of more than 40,000 records.

It was seen that the databases had between 300 and 42,000 records, depending on the surgery, however, all of them suffered from noise, that is, inconsistent data that invalidated certain analyses. During the cleaning process, there were losses of 10 to 60% of records from the original database. The algorithms that were verified followed the criterion:

- Variation of the type of MLP algorithms: 5 types;
- Variation in the number of neurons in the hidden layer: 3, 5, 7;
- Variation in the resampling percentage: 65/35% and 75/25%;

Thus, the students performed the tests in 5 rounds and acquired the average AUC of each one, totaling 150 experiments per database. The results, in terms of AUC, ranged from undefined (some configuration combinations were not sufficient to understand the database) to accuracies of 98%. The relevance and quality of the writing resulted in more than 20 publications in renowned international conferences, as shown in Table 4. There were more than 180 students, freshmen with impactful publications in the proceedings of renowned international conferences, many of them associated with supplements of high-impact journals (Qualis A1 journals and high-impact journals - greater than 2).

Table 4: Database and Resulting Papers

Groups of 2018/2	Groups of 2019/1	Groups of 2019/2
General - Souza et al. (2019.a), Souza et al. (2020.a)	Knee Arthroplasty - Souza et al. (2020.b)	Aortic Aneurysm - Souza et al. (2020.c)
Cardiac Souza et al. (2019.b),	Hip Arthroplasty - Souza et al. (2020.d)	Colon Surgery - Souza et al. (2020.e)
Plastic Surgery - Souza et al. (2020.f)	Limb Amputation - Souza et al. (2020.g)	Prosthetic Surgery - Souza et al. (2023.a)
Orthopedic Surgery - Souza et al. (2019.c),	Bariatric Surgery - Souza et al. (2020.h)	Emergency Surgery - Souza et al. (2023.b)
Craniotomy - Souza et al. (2019.d),	Cholecystectomy - Souza et al. (2021.a)	Infected Surgeries - Souza et al. (2020.i)
Vascular - Souza et al. (2019.e),	Pacemaker Implantation - Souza et al. (2021.b)	Clean Surgeries - Souza et al. (2020.j)
Obstetrics - Souza et al. (2019.f)	Abdominal Hysterectomy - Souza et al. (2020.k)	Small Intestine - Souza et al. (2023.b)
-	Revascularization - Souza et al. (2020.l)	Exploratory Laparotomy - Souza et al. (2023.b)
-	-	Patients Over 70 Years Old Souza et al. (2020.m)
-	-	Orthopedic Trauma - Souza et al. (2023.a)

Source: (Authors)

In discussions with students, the learning process for conducting the experiments was considered "challenging." However, the engagement of the 2019/1 and 2019/2 classes was much greater, influenced by the results of the 2018/2 class. An example of results from one of the published articles, referring to the proposed process, is that of Souza et al. (2020.g), described as: "*From 969 data, only 507 were intact for analysis. Statistically: in 12.45% there was an incidence of global infection and that in 10.67% of the cases were SSI (among which, 94.6% had to be hospitalized for more than 10 days); patients were hospitalized on average 21 days (from 0 to 141 days); the average duration is 78 minutes (maximum 360 minutes); 53 deaths (at 16.98% death rate in case of SSI). A maximum prediction power of 0.688 was found.*".

3.2 Selection Bias, Generalizability, Authorship Ethics

The nature of this case study imposes limitations. Selection bias and generalizability are confirmed, as the data come from a single university center (UNIBH) and partner hospitals within a specific time frame (2018/2, 2019/1, and 2019/2), limiting the extrapolation of pedagogical results to other institutions or medical curricula.

Furthermore, the significant data loss (10% to 60% of the original records) during the cleaning stage introduces a potential selection bias in the Machine Learning (ML) analysis, since the exclusion of records with "noise" may have altered the characteristics of the studied population, compromising the accuracy of the accuracies found (AUCs).

Regarding authorship ethics, rigor demands that the inclusion of all 6-7 students as co-authors be explicitly justified, confirming that each one fulfilled the authorship criteria of the International Committee of Medical Journal Editors (ICMJE), notably through contribution to the design of the project, data acquisition, analysis, interpretation, or manuscript writing and revision. The students in this project focused on: analysis, interpretation, and manuscript writing.

3.3 Educational Performance

It was duly demonstrated that the connection between medical informatics studies, data analysis in practice, and the use of algorithms and tools that play the role of infection prediction has brought positive results to the academic field itself. But it has also brought a differentiated experience for the students who, even as freshmen, learned and appropriated new techniques and technologies.

The projects validate some of the university's roles: relevant contribution to quality medical science, studies that have the potential to improve professional practice, and the acquisition of new skills by academics. This also stems from the fact that it is a proposal that meets the demands and transformations of society and science, which are becoming increasingly connected with AI technologies. Therefore, the positive collaboration of projects that use AI as a form of analysis or prediction in the area of medical education is highlighted.

This achievement is extremely relevant, as this knowledge transcends the walls of the university in three instances: individually, professionally, and socially.

Individually, students benefited from the experience in the field of science and from the practical experimentation with the use of learning algorithms. In addition, these students, even in their first semester of college, were empowered to face the challenges of experimenting with, manipulating, and managing real and practical data. Tools that were previously very much geared towards other sectors are now embedded in medical tools. This is a unique experience for a freshman student and goes far beyond the historically traditional medical education.

Professionally, the methodology with AI data analysis has provided different beneficial impacts, generating several internationally relevant works, such that congresses associated with Qualis A1 journals and high-impact journals were able to confirm, in a more conventional way, the relevance of the studies.

Socially, due to its relevant significance, the medical informatics teaching method is a tool capable of optimizing services, management, and hospital quality, especially when projecting a long-term scenario. Once medical doctors graduate with the skills to also use AI technologies as a means of collecting and analyzing highly complex data, there is potential for optimizing these healthcare services.

Table 5 shows the results of the pre-experiment and post-experiment assessments regarding the students' familiarity with the study's requirements. Adding this to the learning assessment objectives, it can be stated that:

Table 5: Pre/Post Measures-assessment of familiarity with the resources of the experiments

Evaluation	Reference	Yes (rate)
Pre	Are you proficient with computers? (yes or no)	15%
Pre	Do you use or have you used Microsoft Excel or Google Sheets? (yes or no)	35%
Pre	Do you have any basic knowledge of programming? (yes or no)	3%
Pre	Are you comfortable with work involving computers? (yes or no)	52%
Pre	Do you understand that data analysis is relevant to your profession? (yes or no)	80%
Pre	Are you familiar with scientific articles? (yes or no)	7%

Evaluation	Reference	Yes (rate)
Post	Did you effectively participate in the computational part? (yes or no)	18%
Post	Did you understand the meaning of the data in your database? (yes or no)	87%
Post	Did you understand the results of the MLP algorithm? (yes or no)	87%
Post	Have you ever attended or are you familiar with scientific conferences? (Yes or No)	23%

Source: (Authors)

Regarding the understanding of the data to be worked with, it was verified that understanding reached approximately 87%. In this case, the students minimally understood: what a database is, what the data meant, and how important it was for strategic decisions. The post-validation element reports, it was requested that data specify which variables were removed and how many records remained, with consistency exceeding 80% proximity to a clean dataset by 92% of the teams (2 teams required revaluation), and the presentation of at least one coherent justification of the accuracy rate (AUC) versus the imbalance of the dataset. Validation was provided through the resulting article, which also served as the basis for the published articles.

As for the verification of technical skills, in demonstrating technical proficiency in manipulating the R® software to apply the Multi Layer Perceptron (MLP) algorithm in the prediction of Surgical Site Infection (SSI), it was noted that in 100% of all groups from all semesters, between 1 and 2 students in each of the connected groups demonstrated all the necessary capabilities of the experiment, such as: installing the RStudio® software (without teacher support) on their computers; the students were able to receive, open, and manipulate the database using tools like Excel, and then use the manipulated file to populate the MLP running in RStudio®. The remaining students did not demonstrate satisfactory knowledge of software installation capabilities. It can be seen that the ability to manipulate data did not become effective with the students due to difficulties with the tool. However, reading and interpretation became significantly effective.

In the demonstration of critical thinking, in the final report, in the format of a scientific article, a significant difference was noted compared to the previous analysis. While in the computational technical aspect, 1 to 2 students had engaged in the experimentation process, all students had engaged, with at least 80% knowledge in the evaluation of the articles, regarding the evaluation of the results obtained. They were able to distinguish whether the AUC was valid or not, which MLP mode had the best result, and what the most striking characteristics of the database were, such as average age, death rate, SSI rate, and other consolidated data.

Finally, in addition to the publications presented, a semester performance evaluation can be seen in Table 6.

Table 6: Team ratings

Semester	Number of teams	Mean	Min	Max	Std. Deviation	Variance
2018/2	7	45,135	42,5	46,8	1,32	1,81
2019/1	8	49,24	47,67	50	0,67	0,461
2019/2	10	47,81	43,5	49,5	1,10	1,47

Source: (Authors)

The table shows the lowest average grade in the first semester, when the project was proposed. The engagement, since the proposal was considered audacious, was still low compared to the others. Also in this semester, the lowest of the lowest grades was obtained. However, there is greater and more consistent engagement in the others, reaching grades close to or exactly at 50. It can be deduced that the success of the publications from the initial classes engaged the subsequent classes. The program's effectiveness should therefore be analyzed not only by scientific output (20 publications), but also by the lasting acquisition of competencies. The practical integration of machine learning algorithms addresses the growing demand for Data Literacy competencies in the medical curriculum, considered essential for future clinical decision-making (Hoffmann et al., 2022). By manipulating noisy data and applying the Multi Layer Perceptron (MLP) algorithm in the R® software, the student not only learns the definition of AI, but experiences it as an active diagnostic and management tool (Lee et al., 2021). This hands-on experience, which is "challenging" for the students, is crucial for the development of technical proficiency and analytical reasoning, results that require impact assessments in the technology domain to be fully validated.

4 Conclusion

It is concluded that the proposal provided a unique experience for medical students. This experience, even with the difficulties of practical experience and experimentation, as reported by them, brought skills that are increasingly demanded by the job market and society, but which are still little explored in medical courses.

Despite being freshmen in academia, these groups managed to produce high-impact publications in international journals and conferences, providing traditional support for the project. This achievement at such an early stage of the course provides potential mastery for other high-impact and large-scale publications, contributing to the advancement of medical science through technology.

Thus, given so many positive learnings and impacts, which extend beyond the university environment, the medical students accessement to the development of data connection and analysis skills using AI algorithms. To this end, more research, publications, and projects like these should be encouraged.

References

Cruz, K. R., da Silva Toledo, R., de Oliveira, A. S., Moreira, A. M. and Gandin, L. R. A. (2023). AI in the classroom: how Artificial Intelligence is redefining teaching methods [*IA na sala de aula: como a Inteligência Artificial está redefinindo os métodos de ensino*]. Rebena-Revista Brasileira de Ensino e Aprendizagem [*Brazilian Journal of Teaching and Learning*], 7, 19-25.. Available in: https://rebena.emnuvens.com.br/revista/article/view/128

Gansemer-Topf, A. M., McCloud, L. I. and Braxton, J. M. (2024). Defining the scholarship of teaching and learning (SoTL). New Directions for Student Services, 2024(185), 9-17. https://doi.org/10.1002/ss.20502

Kokotsaki, D., Menzies, V. and Wiggins, A. (2016). Project-based learning: A review of the literature. Improving schools, 19(3), 267-277. https://doi.org/10.1177/1365480216659733

Lee, J., Wu, A. S., Li, D. and Kulasegaram, K. (2021). Artificial intelligence in undergraduate medical education: a scoping review. Academic Medicine, 96(Supplement_2), S62-S70. https://doi.org/10.1097/ACM.0000000000004291

Hutchings, P. and Shulman, L. S. (1999). The scholarship of teaching: New elaborations, new developments. Change: The Magazine of Higher Learning, 31(5), 10–15.. https://doi.org/10.1080/00091389909604218

Hoffmann, I., Behrends, M., Highmeducation Consortium and Marschollek, M. (2022). Data Literacy in Medical Education-An Expedition into the World of Medical Data. In ICIMTH (pp. 257-260). https://doi.org/10.3233/SHTI220711

Ricieri, D. V., de Farias, A. M. G., Barreto, R. V. G. and de Souza, F. R. (2024). Common mistakes made by teachers lacking literacy in Artificial Intelligence: an integrative review for higher education [Erros comuns de docentes sem letramento em Inteligência Artificial: uma revisão integrativa para o ensino superior], Peer Review, 6(7), 284-300. https://doi.org/10.53660/PRW-1986-3703

Souza F.H.B., Couto, B.R.G.M., Braga, G.M., Teixeira, J.A., Santos R.C., Martins J.M.C., et al. (2019.a) Medicine allied to technology: The use of artificial neural networks in the prediction of surgical site infection for general surgery service. Abstracts from the 5th International Conference on Prevention & Infection Control; set 09; Geneva. Switzerland. Geneva: ARIC, 2019, 8 (148 Suppl 1). P33.

Souza, F. and Mourão, P. (2020). The Use of Pattern Recognition Techniques in the Detection of Attack Patterns in IDS [*O Uso de Técnicas de Reconhecimento de Padrões na Detecção de Padrões de Ataque em IDS*]. Electronic Journal of Scientific Initiation in Computing [*Revista Eletrônica de Iniciação Científica em Computação*], 18(4). Available: https://journals-sol.sbc.org.br/index.php/reic/article/view/1770

Souza, F. H. B. D. S., Couto, B. R. G. M., Maroca, A. M. R., Soares, I. H. S., Duarte, J. M. D. C. and Costa, S. C. (2019.f). Power of prediction of surgical site infection in caesarean surgery using pattern-based analysis based on multilayer perceptron artificial neural networks. Antimicrobial Resistance & Infection Control, 8(148 Suppl 1). P32.

Souza, F. H. B. D. S., Couto, B. R. G. M., Mello, F. B., Caldeira, I., Morais, J. S. F. R., Mota, L. G., et al. (2019.e). Algorithms for pattern recognition for prediction of surgical site infection in vascular surgeries. Antimicrobial Resistance & Infection Control, 8(148 Suppl 1). P36.

Souza, F. H. B. D. S., Couto, B. R. G. M., Michelini, A. C. L., Melo, A. P. R., Guerra, L. G., Costa, L. M. B. et al. (2019.b). Risk of surgical site infection after cardiac surgeries: Pattern-based analysis based on multilayer perceptron artificial neural networks. Antimicrobial Resistance & Infection Control, 8(148 Suppl 1). P35.

Souza, F. H. B. D. S., Couto, B. R. G. M., Serpa, G. S. D. M., Abelha, I. P. D. A. L., Valadao, L. F., Bernardes, M. E. C. et al. (2019.c). Using artificial neural networks to predict surgical site infection in orthopedic patients. Antimicrobial Resistance & Infection Control, 8(148 Suppl 1). P34.

Souza, F. H. B. D., Couto, B. R. G. M., Matias, J. D. O., Araujo, L. L., Polidoro, L. R., & Rossati, L. S. (2019.d). Risk prediction for surgical site infection in craniotomy patients. Antimicrobial Resistance & Infection Control, 8(148 Suppl 1). P31.

Souza, F. H. B., Couto, B. R. G. M., Conceição, F. L. A., Martins, M. B., Pimenta, G. M., Mendes, J. C. O. et al. (2023.b). Surgical site infection predictive analysis: A comparative evaluation of multiple surgeries procedures. Antimicrobial Resistance & Infection Control, 12(81 Suppl 1). P30.

Souza, F. H. B., Couto, B. R. G. M., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2020.b). Risk prediction for surgical site infection in patients subject to knee arthroplasty surgery. Open Forum Infectious Diseases, 7(Suppl 1), S484–S485. https://doi.org/10.1093/ofid/ofaa439.1089

Souza, F. H. B., Couto, B. R. G. M., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2020.c). Artificial neural networks to predict surgical site infection in aorta artery aneurysm correction. Open Forum Infectious Diseases, 7(Suppl 1), S476–S477. https://doi.org/10.1093/ofid/ofaa439.1068

Souza, F. H. B., Couto, B. R. G. M., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2020.e). Prediction of surgical site infections in colon surgery in Belo Horizonte hospitals. Open Forum Infectious Diseases, 7(Suppl 1), S483. https://doi.org/10.1093/ofid/ofaa439.1085

Souza, F. H. B., Couto, B. R. G. M., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2020.g). Occurrence's prediction of surgical site infection in limb amputation surgery. Open Forum Infectious Diseases, 7(Suppl 1), S481. https://doi.org/10.1093/ofid/ofaa439.1081

Souza, F. H. B., Couto, B. R. G. M., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2020.h). Prediction of surgical site infection risk in patients undergoing bariatric surgery. Open Forum Infectious Diseases, 7(Suppl 1), S482–S483. https://doi.org/10.1093/ofid/ofaa439.1084

Souza, F. H. B., Couto, B. R. G. M., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2020.i). Prediction of occurrence for surgical site infection in infected surgeries. Open Forum Infectious Diseases, 7(Suppl 1), S482. https://doi.org/10.1093/ofid/ofaa439.1083

Souza, F. H. B., Couto, B. R. G. M., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2020.j). Technology and medicine: Prediction of surgical site infection in clean surgeries using artificial neural networks. Open Forum Infectious Diseases, 7(Suppl 1), S487. https://doi.org/10.1093/ofid/ofaa439.1095

Souza, F. H. B., Couto, B. R. G. M., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2020.m). Artificial neural networks to predict infection in the surgical site in patients over 70 years old. Open Forum Infectious Diseases, 7(Suppl 1), S476. https://doi.org/10.1093/ofid/ofaa439.1067

Souza, F. H. B., Couto, B. R. G. M., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2021.a). Artificial neural networks and medicine: An interdisciplinary approach to surgical site infection in cholecystectomy. Proceeding of 31st European Congress of Clinical Microbiology and Infectious Diseases (ECCMID), Viena, Áustria.

Souza, F. H. B., Couto, B. R. G. M., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2021.b). Cardiac pacemaker implantation surgery: Automated prediction of surgical site infection. Open Forum Infectious Diseases, 8(Suppl 1), S499. https://doi.org/10.1093/ofid/ofab466.1006

Souza, F. H. B., Couto, B. R. G. M., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2023.a). Prediction of surgical site infection: A comparative analysis between two surgeries based on artificial neural networks. Antimicrobial Resistance & Infection Control, 12(81 Suppl 1). P29.

Souza, F., Couto, B., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2020.d). Automated risk analysis of surgical site infection in hip arthroplasty surgeries. Infection Control & Hospital Epidemiology, 41(Suppl 1), S135–S136. https://doi.org/10.1017/ice.2020.649

Souza, F., Couto, B., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2020.f). Artificial Neural Networks Applied to Prediction to Assess the Likelihood of Surgical Site Infection in Different Surgeries. Infection Control & Hospital Epidemiology, 41(S1), s129–s129. https://doi.org/10.1017/ice.2020.639

Souza, F., Couto, B., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2020.k). Pattern recognition algorithms for predicting surgical site infection in abdominal hysterectomy. Infection Control & Hospital Epidemiology, 41(Suppl 1), S344–S345. https://doi.org/10.1017/ice.2020.958

Souza, F., Couto, B., Conceição, F. L. A., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M. et al. (2020.l). Automated prediction of surgical site infection coronary artery bypass (CABG) grafting surgery. Infection Control & Hospital Epidemiology, 41(Suppl 1), S135. https://doi.org/10.1017/ice.2020.648

Souza, F., Couto, B., Silva, G. H. S., Dias, I. G., Rigueira, R. V. M., Pimenta, G. M. et al. (2020.a). Prediction of surgical risk in general surgeries: Process optimization through Support Vector Machine (SVM) algorithm. Infection Control & Hospital Epidemiology, 41(Suppl 1), S355–S356. https://doi.org/10.1017/ice.2020.975

Thomas, J. W. (2000). A review of research on project-based learning. The Autodesk Foundation.

Vicari, R. M. (2021). The influence of artificial intelligence technologies on education [*Influências das tecnologias da inteligência artificial no ensino*]. Advanced Studies [*Estudos Avançados*], 35(101), 73–84. https://doi.org/10.1590/s0103-4014.2021.35101.006