Improving management of febrile neutropenia in oncology patients: the role of artificial intelligence and machine learning

Journal:	*Expert Review of Anti-infective Therapy*
Manuscript ID	ERI-2023-ST-0214.R1
Manuscript Type:	Review
Keywords:	Febrile Neutropenia, Artificial Intelligence, Machine Learning, Supervised Learning, Personalised Medicine, Clinical Decision-making, Bloodstream infections, Multidrug-Resistant Bacteria

AI and ML in febrile neutropenia

ABSTRACT

Introduction: Artificial intelligence (AI) and machine learning (ML) have the potential to revolutionise the management of febrile neutropenia (FN) and drive progress towards personalised medicine.

Areas covered: In this review, we detail how the collection of a large number of high-quality data can be used to conduct precise mathematical studies with ML and AI. We explain the foundations of these techniques, covering the fundamentals of supervised and unsupervised learning, as well as the most important challenges, e.g., data quality, "black box" model interpretation and overfitting. To conclude, we provide detailed examples of how AI and ML have been used to enhance predictions of chemotherapy-induced FN, detection of bloodstream infections (BSIs) and multidrug-resistant (MDR) bacteria, and anticipation of severe complications and mortality.

Expert opinion: Authors' expert opinion focused on the promising potential of implementing accurate AI and ML models whilst managing FN. However, their integration as viable clinical tools poses challenges, including technical and implementation barriers. Improving global accessibility, fostering interdisciplinary collaboration and addressing ethical and security considerations are essential. By overcoming these challenges, we could transform personalised care for patients with FN.

KEYWORDS

Febrile Neutropenia, Artificial Intelligence, Machine Learning, Supervised Learning, Personalised Medicine, Clinical Decision-making, Bloodstream Infections, Multidrug-Resistant Bacteria

AI and ML in febrile neutropenia

ARTICLE HIGHLIGHTS

1. **AI in healthcare:** AI and ML can improve the management of FN in patients. Their adoption will depend on a better understanding of how they work.
2. **Data quality in AI:** Managing large volumes of high-quality data and interdisciplinary work are key to the success of predictive studies in AI. Tools like SQL, Python, and R are crucial.
3. **AI in clinical practice:** The integration of AI faces challenges such as the perception of 'black boxes', the need for more transparency in algorithms, and the risk of overfitting.
4. **ML and assistance for neutropenic patients:** ML models outperform traditional methods in predicting complications in patients with FN. At the same time, they can minimise subjectivity in clinical decisions, overcoming the limitations of current clinical scales.
5. **Challenges and obstacles:** Despite the potential of AI and ML to revolutionise the management of FN, there are barriers to overcome, including the availability and management of high-quality data; the integration of real-time data collection and processing; and the complete collaboration amongst technical and clinical teams.

AI and ML in febrile neutropenia

1. Introduction

In the contemporary landscape of healthcare, personalised medicine heralds a paradigm shift towards an individual-centric approach to care [1]. The synergy of data-driven methodologies, especially machine learning (ML), and the broader artificial intelligence (AI) spectrum power this transition [2–8]. Indeed, this reality is already making its mark in the management of febrile neutropenia (FN), a common and potentially life-threatening complication in patients undergoing myelosuppressive chemotherapy. Although the application of ML in this field may appear futuristic, it is a tangible existence within present times. Throughout this comprehensive review, our objective was to elucidate the contributions of personalised medicine and the statistical methodologies to be employed, and provide clinicians with a concise explanation regarding ML algorithms. We aimed to underscore the transformative potential of ML in enhancing clinical decision-making processes and administering more efficacious and personalised treatments for patients with FN.

2. What is the initial step in enabling artificial intelligence? Data quality.

Historically, physicians have arrived to decisions based on studies relying on manual data entry with a relatively low number of patients and variables. The revolution of AI and ML starts with their ability to handle a high volume set of data, which could perhaps best be retrieved automatically from electronic health records (EHRs) or multiple interconnected devices [2,5,9–13]. Whilst automatic collection is not foolproof against mistakes, rigorous oversight of data quality remains crucial. This essential notion is embodied in the adage often cited in data science, 'garbage in, garbage out', which

stresses the importance of data quality in defining the accuracy and reliability of conclusions derived from these models [14].

The use of programming languages such as Structured Query Language (SQL), Python (Python Software Foundation, Wilmington, DE), and R (R Foundation for Statistical Computing, Vienna, Austria) is essential nowadays for managing databases; manipulating, visualising and assessing data quality; and constructing those models [15–17]. These tools help address challenges associated with EHR storage and access, and foster the creation of comprehensive datasets. This necessitates a multidisciplinary approach involving collaboration amongst physicians, bioinformatics experts and data-scientists.

The ability of such technology to handle vast amounts of data enables the possibility of conducting predictive studies with a different perspective in comparison to those prior. For example, in the field of FN, most published studies typically focus only on patients with confirmed infections. Therefore, applying any outcome to FN onset, when physicians do not known whether the culture will be positive or not, represents a significant initial bias [6]. Thanks to this capability in analysing a large volume of high-quality data, two recent studies have trained data to predict whether a neutropenic patient will have a positive culture and/or a positive culture due to multidrug-resistant (MDR) bacteria at FN onset [18,19].

Another significant advantage of these models is their capacity to collect data continuously, thereby ideally enhancing predictive accuracy over time. For instance,

constant data generation from wearable devices monitoring physiologic metrics, such as body temperature, has been used to detect fever episodes in high-risk patients, e.g., those undergoing haematopoietic stem cell transplantations (HSCT). One interesting study showed the effective use of non-linear autoregressive models with exogenous inputs (NARX), a type of artificial neural network (ANN) algorithm, in detecting clinically-assessed fever episodes with high sensitivity (90.2%) and specificity (87.8%) [20]. This early detection approach may help in the prompt identification and early treatment of FN, thereby improving clinical outcomes [21,22].

Another example that showed how this integration may improve clinical decision-making processes is the use of this technology in interpreting chest x-rays (CXR) ~~common diagnostic tool for pneumonia that plays a crucial role in the initial investigation in patients with FN~~ [23,24]. However, as compromised immune responses can impede the timely development of radiographic infiltrates, challenges can arise when it comes to early, accurate detection of pneumonia via CXR [25]. In a study conducted by Hwang *et al.* ~~[26]~~, authors developed a sophisticated computer-aided detection (CAD) system, driven by deep learning algorithms based on ANN [26]. This investigation used a robust dataset from the general population, comprising 54,220 normal CRXs and 35,613 abnormal CXRs, including 6,903 from patients with pneumonia. In another related study with a total cohort of 413 patients with FN, the same authors showed that the integration of CAD significantly enhanced physicians' diagnostic accuracy of pneumonia from 75.4% to 79.5% [27].

3. Is it plausible for physicians to transcend the 'Black Box' obstacle? The path towards understanding and effective implementation in clinical practice.

Notwithstanding promising results, the full potential of research in this field has not completely translated into clinical practice for various reasons. A prevailing sentiment—challenging to alleviate, though—is the unease associated with replacing clinical decision-making processes with opaque decision tools, commonly referred to as 'black boxes', and possible consequences of computational errors [13,16,28–30]. Many studies on ML have trouble in communicating their methodologies effectively: algorithms employed are not disclosed, and the datasets used are not revealed. This, as a result, have left readers with limited opportunities to thoroughly scrutinise results for inaccuracies [31]. To promote result reproducibility and build confidence in these techniques, a comprehensive review and basic explanation of the statistical approach are necessary. The collaboration between data scientists and medical professionals appears to be essential in such a case. In addition to helping better the understanding of these mathematical algorithms, physicians must guide the selection of models needed in clinical practice.

4. When should I use supervised or unsupervised learning, and which algorithm is better suited to answer the question concerning my problem? Basic concepts.

To equip physicians with a stronger understanding of how these complex ML algorithms work, it is essential to approach the learning problem more broadly. ML techniques may be generally classified as supervised or unsupervised learning ~~(**Figure 1**)~~. Supervised learning involves training an algorithm using labeled input data, with the aim of

predicting outcomes (outputs) for new unlabeled data (Figure 1). This approach is mainly used in both classification and regressions tasks: in the former, categories are assigned to the data whilst in the latter, a continuous numerical value is predicted. On the other hand, unsupervised learning involves training a model with unlabeled data, thereby allowing the model to discover patterns and relationships in the data without explicit guidance (Figure 2) [2,32,33]. Supervised learning can be said to be capable of performing both classification and regression/prediction tasks, whilst unsupervised learning, clustering tasks. To note, the specific application of unsupervised learning algorithms, such as K-means or hierarchical clustering, has not yet been explored in the context of FN.

Each ML algorithm has its own strengths and weaknesses in terms of the types of input data, performance and interpretation. **Table 1** provides a summary of the main categories of structured algorithms used, including an intuitive description and their key characteristics.

Understanding the performance metrics employed to assess the quality of ML models is nearly as vital as comprehending the mathematical mechanisms of the algorithms. Common metrics encompass accuracy, precision and recall [12,13,46]. The F1 score is a metric that combines both the positive predictive value (precision) and true positive rate (recall or sensitivity). Unlike precision and recall separately, it considers both false positives and false negatives simultaneously. However, it still omits true negatives. It is typically used instead of accuracy in cases of severe class imbalance in the dataset [47]. Area under the receiver operating characteristic (AUROC), a widely

employed numerical metric, serves to appraise the overall performance of the model in binary classification tasks [2]. In implementing these algorithms within clinical practice, it is paramount not to excessively focus on the novelty of the technique, but instead, holistically grasp the prognosticated risk's nature, its precision level and the context in which this model has been validated.

5. What is overfitting? The great enemy of AI algorithms.

As previously mentioned, AI algorithms use a dataset for training with the objective of identifying key features and adjusting parameters. A common problem is when the model memorises the data points from the training set and overly fits to them, almost like a photograph. If it does this, it is not learning "the problem" but rather "memorising points". Consequently, when trying to classify new unlabelled data, the algorithm loses precision or simply fails [31,48,49]. This problem is called overfitting. In response to this challenge, different approaches can be employed, including cross-validation. This strategy repeats the division of data into training sets multiple times, using different combinations of data in each iteration. The goal of cross-validation is to evaluate the model's performance in an unbiased and robust manner, so as to reduce the risk of overfitting and improve its ability to generalise for a broader population [44].

6. Are ML approaches available nowadays to help neutropenic patients? What are the most important differences with non-ML studies? Present, not future.

Physicians' clinical decision-making processes are based on attempts to predict what patients need at a specific moment. For this reason, medical research has generated multiple prediction scales over the years. The most common, current clinical scores,

such as the Multinational Association for Supportive Care in Cancer (MASCC) risk index or the Clinical Index of Stable Febrile Neutropenia (CISNE), rely on clinical and laboratory criteria, hospitalisation status, presence of comorbidities and underlying malignancy type to risk-stratify patients at the onset of FN [50,51]. However, these scales have presented some limitations such as focusing on very specific populations, lacking external validations, or depending heavily on hospital-specific data.

Table 2 details the current published ML approaches to support clinical decision-making processes in FN. A recent investigation by Hui *et al*. ~~[43]~~ assessed the ability of an ANN model to improve the outcome prediction of chemotherapy-induced FN in patients with onco-haematological malignancies [43]. The authors used a cohort of 227 patients. The ANN model used 14 parameters, including both the criteria from the MASCC risk-index, and some variables identified by their univariable analysis. The ANN model could not meet the performance of the MASCC risk-index in this population, with an AUROC of 0.74 vs 0.81. When contrasting this model, which failed to demonstrate a significant improvement over the original MASCC score, Padmanabhan *et al.* ~~[18]~~ conducted a study aimed at predicting three critical clinical complications—sepsis, the presence of MDR organisms and mortality—from the data made available in medical records including 1166 episodes of FN [18]. The model proposed demonstrated enhanced recall and AUROC in predicting sepsis (recall = 98%, AUROC = 0.85), MDR organism (recall = 96%, AUROC = 0.91), and mortality (recall = 86%, AUROC = 0.88). Another study conducted by Venäläienen et *al.* ~~[52]~~ aimed to predict the risk of neutropenic infection with a penalised regression model [52]. For these authors, neutropenic infection was defined as grade IV neutropenia with serum C-reactive protein >10 mg/L. The model

demonstrated good precision, with an AUROC of 0.84. During the external validation phase, it outperformed other models based on traditional statistical approaches and achieved an AUROC of 0.75.

Recently, Du *et al.* [39] demonstrated that it is feasible to accurately identify hospitalised adults with FN who have a high risk of mortality during hospitalisation whilst avoiding subjective variables [39]. This finding is contrary to the MASCC risk-index, which encompasses the subjective assessment of the burden of illness determined by the attending physician at patient presentation. To achieve this objective, they also used a non-linear model, i.e., gradient boosting machine (GBM), yielding substantial enhancements in prediction, with AUROC of 0.92 and F1 score of 0.75.

Another highlight is the identification of patients at high risk of bloodstream infection (BSI). This has become a priority focus in predictive models due to its association with high morbidity and mortality [54–56]. A recent study in a paediatric population harnessed EHRs to craft an ML model aimed at predicting BSI. This algorithm has been applied to over 11,000 blood cultures drawn from oncologic patients or HSCT recipients. In this study by Sung *et al.* [37], an AUROC of 0.74 was achieved using a GBM model [37]. However, the performance of the algorithm did not significantly exceed the BSI prediction based solely on the presence of neutropenia. On the other hand, the study by Alali *et al.* [38] used a different approach using a random forest (RF) model and demonstrated higher accuracy in predicting BSI, with an AUROC of 0.79 [38].

7. Conclusion

AI and ML in febrile neutropenia

The field of AI and ML holds significant potential to improve the management of FN. However, to fully leverage these tools, we need to tackle key challenges related to data availability, the understanding of performance metrics, effective integration of data collection and processing, and implementation of real-time application in hospital systems. By moving past these hurdles, we can bolster the accuracy of our predictions, personalise healthcare, and ultimately improve patient outcomes.

EXPERT OPINION

The development of accurate and reliable ML models may be extremely promising in the management of FN. However, the mere creation of these models is not enough. The ultimate aspiration is to transform them into clinically applicable tools. The main challenge lies in bridging the gap between computational results and everyday clinical decision-making processes. This goal requires overcoming technical barriers and addressing integral aspects of medical practice and healthcare institutions.

To date, several groups have been working on AI and ML models to explore potential enhancements in patient management. The next steps include fostering closer collaboration between clinicians and these models, encouraging clinicians to take on proactive roles as proponents or advocates of these algorithms. Moreover, conducting studies that demonstrate the efficacy and efficiency of AI/ML models in clinical practice and developing tools that facilitate the interaction between physicians and algorithms are essential objectives.

A critical aspect in overcoming these challenges involves establishing specific standards for evaluating data quality. While there are initiatives that provide a framework for the presentation and validation of predictive models in ML, there is a fundamental need to delve deeper into data quality assessment. Establishing standards involves defining rules and regulations to ensure that data is consistently presented in terms of formats and units, that access is controlled, audits are conducted to verify data quality, and information about the data is clear and understandable. The standardization of these aspects is crucial not only for the validity and reproducibility of these models but also for their acceptance and trust within the medical and scientific community. Despite individual efforts to manage data quality, including peer reviews and cross-validations, the absence of a standardized approach continues to pose a considerable challenge. This challenge is further exacerbated by the heterogeneous and often fragmented nature of data sources in the healthcare field.

The progression towards specific data quality standards and collaboration between clinicians sets the stage for the development of global data warehouses. These repositories will be instrumental in refining models for diverse populations, taking into account factors like race, ethnicity, and social status, thus promoting a more inclusive and equitable usage of these technologies in FN management. By ensuring that these models are tested and validated across a spectrum of demographic and clinical scenarios, the efficacy and applicability of these tools can be significantly enhanced.

The integration of ML methodologies into clinical practice offers more than just prognostic predictions. It provides valuable insights for decision-making in key areas

such as antimicrobial stewardship. Fully validated models could be instrumental in optimizing antimicrobial use, reducing the misuse of antimicrobial agents and inappropriate empirical antibiotic therapies. This advancement is crucial in preventing complications like *Clostridium difficile* infections and controlling the spread of MDR bacteria. The incorporation of these advanced methodologies can help improve the precision and effectiveness of therapeutic strategies in a scenario increasingly challenged by bacterial multi-resistance.

However, it is important to highlight that doctors with extensive clinical experience but without previous knowledge of these methodologies may distance themselves from this transformation or even feel overwhelmed. Therefore, our strong recommendation is that the optimal implementation of these processes involves building multidisciplinary teams, requiring close and synergistic collaboration among physicians, bioinformatics experts, and data scientists.

In conclusion, the effective application of AI and ML presents a challenging yet fascinating endeavour. This undertaking encompasses crucial aspects such as data quality management, fostering collaboration among clinicians, and the imperative to address transparency issues within these algorithms. Successfully facing and resolving the current complexities that these mathematical tools represent for physicians may lead to significant, transformative advancements in personalized care for patients with FN.

REFERENCES

AI and ML in febrile neutropenia

[1] Lin B, Wu S. Digital Transformation in Personalized Medicine with Artificial Intelligence and the Internet of Medical Things. OMICS. 2022;26:77–81.

[2] Handelman GS, Kok HK, Chandra RV, et al. eDoctor: machine learning and the future of medicine. J Intern Med. 2018;284:603–19.

[3] Deo RC. Machine Learning in Medicine. Circulation. 2015;132:1920–30.

[4] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25:44–56.

[5] Obermeyer Z, Emanuel EJ. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016;375:1216–9.

[6] Garcia-Vidal C, Sanjuan G, Puerta-Alcalde P, et al. Artificial intelligence to support clinical decision-making processes. EBioMedicine. 2019;46:27–9.

[7] Toh TS, Dondelinger F, Wang D. Looking beyond the hype: Applied AI and machine learning in translational medicine. EBioMedicine. 2019;47:607–15.

[8] Cirillo D, Valencia A. Big data analytics for personalized medicine. Curr Opin Biotechnol. 2019;58:161–7.

[9] López Pineda A, Ye Y, Visweswaran S, et al. Comparison of machine learning classifiers for influenza detection from emergency department free-text reports. J Biomed Inform. 2015;58:60–9.

AI and ML in febrile neutropenia

[10] Li BY, Oh J, Young VB, et al. Using Machine Learning and the Electronic Health Record to Predict Complicated Clostridium difficile Infection. Open Forum Infectious Diseases. 2019;6:ofz186.

[11] Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care. 2013;51:S30-7.

[12] Alqaissi EY, Alotaibi FS, Ramzan MS. Modern Machine-Learning Predictive Models for Diagnosing Infectious Diseases. Computational and Mathematical Methods in Medicine. 2022;2022:e6902321.

**** This review highlighted recent research articles on ML algorithms applied to the diagnosis of infectious diseases.**

[13] Peiffer-Smadja N, Rawson TM, Ahmad R, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clinical Microbiology and Infection. 2020;26:584–95.

[14] Teno JM. Garbage in, Garbage out—Words of Caution on Big Data and Machine Learning in Medical Practice. JAMA Health Forum. 2023;4:e230397.

[15] Wagholikar KB, Zelle D, Ainsworth L, et al. Use of automatic SQL generation interface to enhance transparency and validity of health-data analysis. Inform Med Unlocked. 2022;31:100996.

AI and ML in febrile neutropenia

[16] Basu S, Faghmous JH, Doupe P. Machine Learning Methods for Precision Medicine Research Designed to Reduce Health Disparities: A Structured Tutorial. Ethn Dis. 2020;30:217–28.

[17] Leenings R, Winter NR, Plagwitz L, et al. PHOTONAI-A Python API for rapid machine learning model development. PLoS One. 2021;16:e0254062.

[18] Padmanabhan R, Elomri A, Taha RY, et al. Prediction of Multiple Clinical Complications in Cancer Patients to Ensure Hospital Preparedness and Improved Cancer Care. IJERPH. 2022;20:526.

*** This study highlights the necessity to popularize AI-based devices as a means to support clinical decision-making.**

[19] Garcia-Vidal C, Puerta-Alcalde P, Cardozo C, et al. Machine Learning to Assess the Risk of Multidrug-Resistant Gram-Negative Bacilli Infections in Febrile Neutropenic Hematological Patients. Infect Dis Ther. 2021;10:971–83.

*** This study suggests these algorithms may help assess MDR-GNB infection risk at FN onset.**

[20] Verma N, Haji-Abolhassani I, Ganesh S, et al. A Novel Wearable Device for Continuous Temperature Monitoring & Fever Detection. IEEE J Transl Eng Health Med. 2021;9:1–7.

[21] Nesher L, Rolston KVI. The current spectrum of infection in cancer patients with chemotherapy related neutropenia. Infection. 2014;42:5–13.

AI and ML in febrile neutropenia

[22] Zimmer AJ, Freifeld AG. Optimal Management of Neutropenic Fever in Patients With Cancer. J Oncol Pract. 2019;15:19–24.

[23] Chumbita M, Cillóniz C, Puerta-Alcalde P, et al. Can Artificial Intelligence Improve the Management of Pneumonia. J Clin Med. 2020;9:248.

[24] Heussel CP. Importance of pulmonary imaging diagnostics in the management of febrile neutropenic patients: Lung imaging in febrile neutropenia. Mycoses. 2011;54:17–26.

[25] Hwang EJ, Park S, Jin K-N, et al. Development and Validation of a Deep Learning–Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs. JAMA Network Open. 2019;2:e191095.

[26] Hwang EJ, Lee JH, Kim JH, et al. Deep learning computer-aided detection system for pneumonia in febrile neutropenia patients: a diagnostic cohort study. BMC Pulm Med. 2021;21:406.

[27] Castelvecchi D. Can we open the black box of AI? Nature. 2016;538:20–3.

[28] Handelman GS, Kok HK, Chandra RV, et al. Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods. American Journal of Roentgenology. 2019;212:38–43.

[29] McCoy LG, Brenna CTA, Chen SS, et al. Believing in black boxes: machine learning for healthcare does not need explainability to be evidence-based. J Clin Epidemiol. 2022;142:252–7.

AI and ML in febrile neutropenia

[30] Andaur Navarro CL, Damen JAA, Takada T, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ. 2021;375:n2281.

[31] An Q, Rahman S, Zhou J, et al. A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges. Sensors (Basel). 2023;23:4178.

[32] Jovel J, Greiner R. An Introduction to Machine Learning Approaches for Biomedical Research. Front Med (Lausanne). 2021;8:771607.

[33] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. 2nd Edition. New York, NY: Springer; 2009.

[34] Doupe P, Faghmous J, Basu S. Machine Learning for Health Services Researchers. Value in Health. 2019;22:808–15.

[35] Emani S, Swaminathan A, Grobman B, et al. Critically reading machine learning literature in neurosurgery: a reader's guide and checklist for appraising prediction models. Neurosurgical Focus. 2023;54:E3.

[36] Cho B-J, Kim KM, Bilegsaikhan S-E, et al. Machine learning improves the prediction of febrile neutropenia in Korean inpatients undergoing chemotherapy for breast cancer. Sci Rep. 2020;10:14803.

[37] Wiberg H, Yu P, Montanaro P, et al. Prediction of Neutropenic Events in Chemotherapy Patients: A Machine Learning Approach. JCO Clinical Cancer Informatics. 2021;904–11.

[38] Sung L, Corbin C, Steinberg E, et al. Development and utility assessment of a machine learning bloodstream infection classifier in pediatric patients receiving cancer treatments. BMC Cancer. 2020;20:1103.

[39] Alali M, Mayampurath A, Dai Y, et al. A prediction model for bacteremia and transfer to intensive care in pediatric and adolescent cancer patients with febrile neutropenia. Sci Rep. 2022;12:7429.

*** This study demonstrated strong BSI prediction in pediatric FN patients at onset.**

[40] Du X, Min J, Shah CP, et al. Predicting in-hospital mortality of patients with febrile neutropenia using machine learning models. Int J Med Inform. 2020;139:104140.

*** This research constructs ML model that do not require physicians' subjective evaluation for FN mortality risk prediction.**

[41] Xiang L, Wang H, Fan S, et al. Machine Learning for Early Warning of Septic Shock in Children With Hematological Malignancies Accompanied by Fever or Neutropenia: A Single Center Retrospective Study. Front Oncol. 2021;11:678743.

[42] Zhan M, Chen Z-B, Ding C-C, et al. Machine learning to predict high-dose methotrexate-related neutropenia and fever in children with B-cell acute lymphoblastic leukemia. Leuk Lymphoma. 2021;62:2502–13.

[43] Jakob C, Classen A, Stecher M, et al. 2187. Prediction of Patient Outcome During Febrile Neutropenia Despite Anti-infective Treatment Using Machine Learning Algorithms. Open Forum Infectious Diseases. 2019;6:S744.

AI and ML in febrile neutropenia

[44] Hui EP, Leung LKS, Poon TCW, et al. Prediction of outcome in cancer patients with febrile neutropenia: a prospective validation of the Multinational Association for Supportive Care in Cancer risk index in a Chinese population and comparison with the Talcott model and artificial neural network. Support Care Cancer. 2011;19:1625–35.

*** This study pioneers ML application in FN.**

[45] Liu Y, Zhou Y, Wen S, et al. A Strategy on Selecting Performance Metrics for Classifier Evaluation. International Journal of Mobile Computing and Multimedia Communications. 2014;6:20–35.

[46] Dankers FJWM, Traverso A, Wee L, et al. Prediction Modeling Methodology. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of Clinical Data Science. Cham (CH): Springer; 2019.

[47] Kernbach JM, Staartjes VE. Foundations of Machine Learning-Based Clinical Prediction Modeling: Part II-Generalization and Overfitting. Acta Neurochir Suppl. 2022;134:15–21.

[48] Koçak B. Key concepts, common pitfalls, and best practices in artificial intelligence and machine learning: focus on radiomics. Diagn Interv Radiol. 2022;28:450–62.

[49] Klastersky J, Paesmans M, Rubenstein EB, et al. The Multinational Association for Supportive Care in Cancer risk index: A multinational scoring system for identifying low-risk febrile neutropenic cancer patients. J Clin Oncol. 2000;18:3038–51.

AI and ML in febrile neutropenia

[50] Carmona-Bayonas A, Jiménez-Fonseca P, Virizuela Echaburu J, et al. Prediction of serious complications in patients with seemingly stable febrile neutropenia: validation of the Clinical Index of Stable Febrile Neutropenia in a prospective cohort of patients from the FINITE study. J Clin Oncol. 2015;33:465–71.

[51] Venäläinen MS, Heervä E, Hirvonen O, et al. Improved risk prediction of chemotherapy-induced neutropenia—model development and validation with real-world data. Cancer Medicine. 2022;11:654–63.

[52] Lind ML, Mooney SJ, Carone M, et al. Development and Validation of a Machine Learning Model to Estimate Bacterial Sepsis Among Immunocompromised Recipients of Stem Cell Transplant. JAMA Network Open. 2021;4:e214514.

[53] Wisplinghoff H, Seifert H, Wenzel RP, et al. Current Trends in the Epidemiology of Nosocomial Bloodstream Infections in Patients with Hematological Malignancies and Solid Neoplasms in Hospitals in the United States. Clinical Infectious Diseases. 2003;36:1103–10.

[54] Klastersky J, Ameye L, Maertens J, et al. Bacteraemia in febrile neutropenic cancer patients. International Journal of Antimicrobial Agents. 2007;30:51–9.

[55] Chumbita M, Puerta-Alcalde P, Gudiol C, et al. Impact of Empirical Antibiotic Regimens on Mortality in Neutropenic Patients with Bloodstream Infection Presenting with Septic Shock. Antimicrob Agents Chemother. 66:e01744-21.

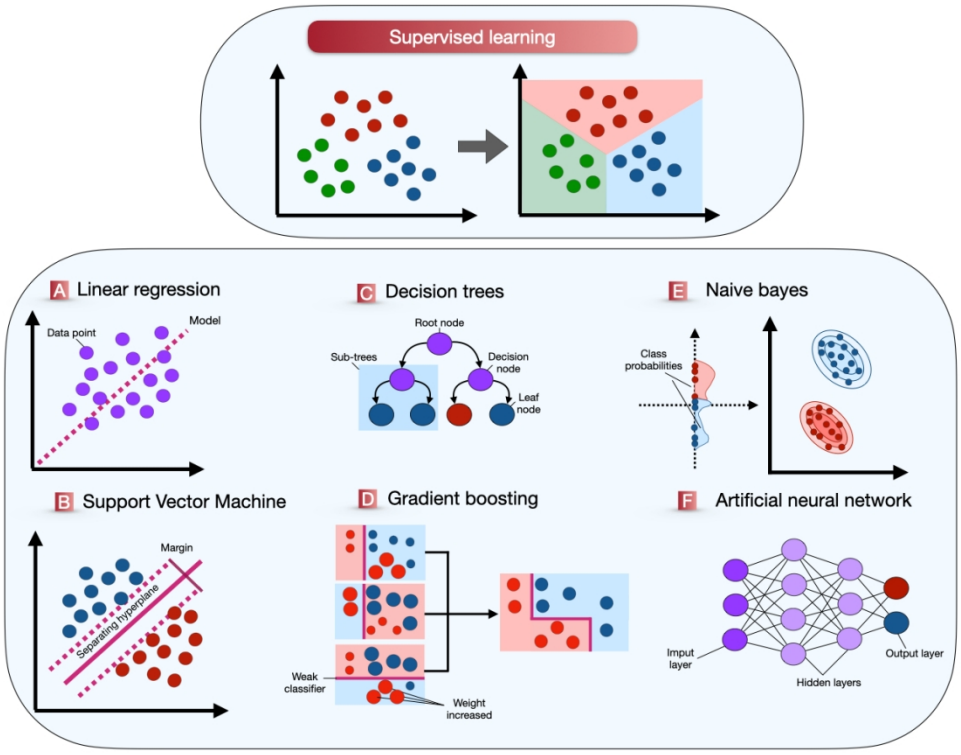


Figure 1. Overview of supervised ML scenario and commonly used algorithms in FN [32–34]
Supervised learning uses labeled data for classification and regression. The bottom panel provide an overview of the prevailing architectural concepts in ML employed to achieve specific objectives: A) Linear regression assumes that the predictor and target variables have a linear relationship; B) Support Vector Machine uses a hyperplane to separate data into different classes, maximising the margin between the decision boundary and the nearest data points; C) Decision trees uses a branching structure to make decisions based on different conditions or attributes of the dataset, with root nodes representing the most important attribute, decision nodes evaluating conditions, and leaf nodes representing final classifications or predictions; D) Gradient boosting combines weak classifiers iteratively to enhance the predictive performance of a model; E) Naive Bayes assigns instances to classes using class probabilities, assuming conditional independence of attributes; F) Artificial neural network learns hierarchical representations of the data through hidden layers, and the output layer performs the final classification.

112x88mm (300 x 300 DPI)

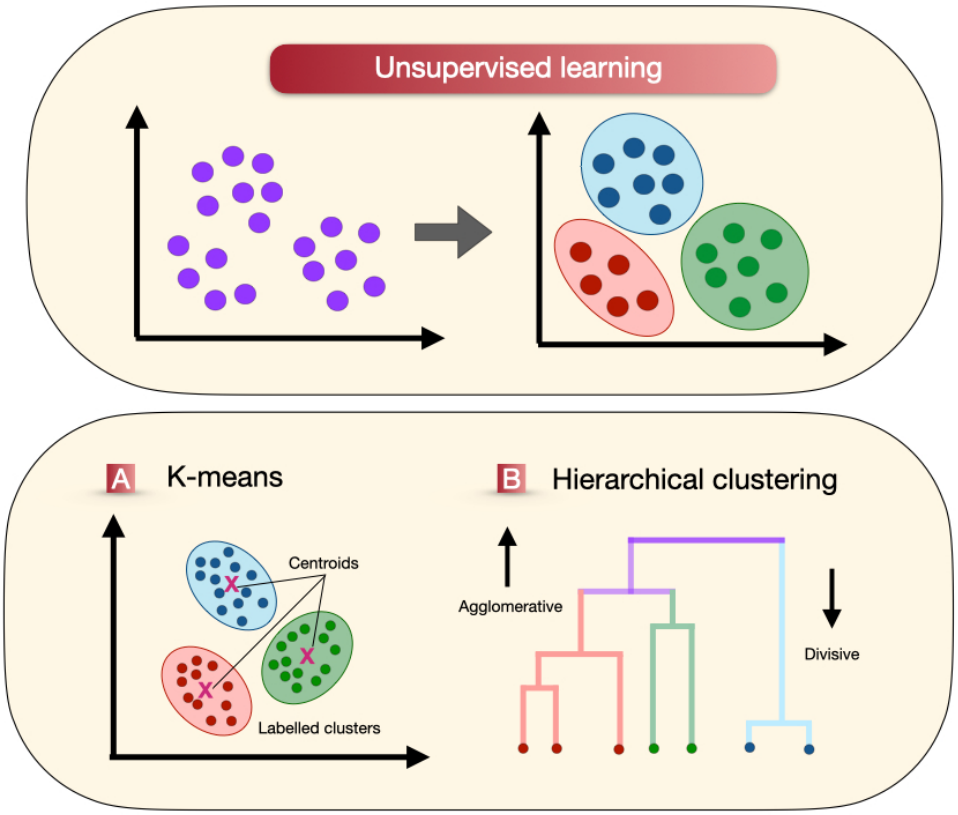
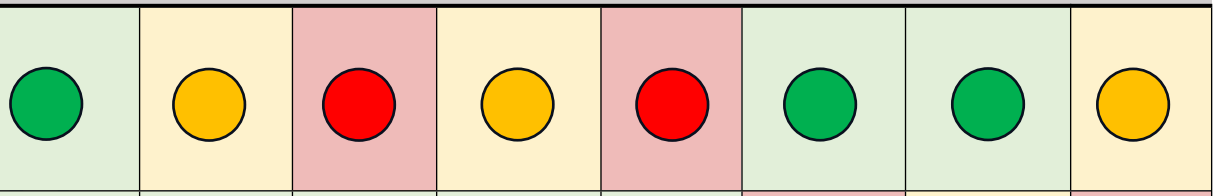
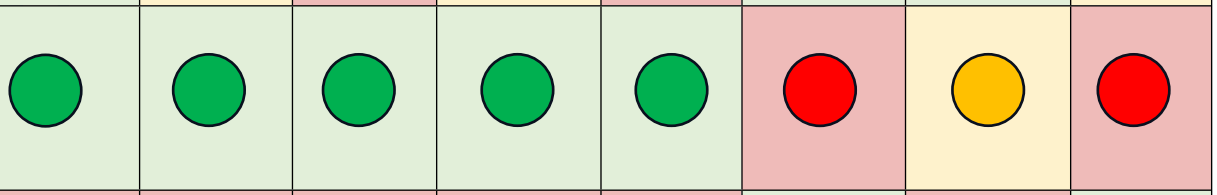
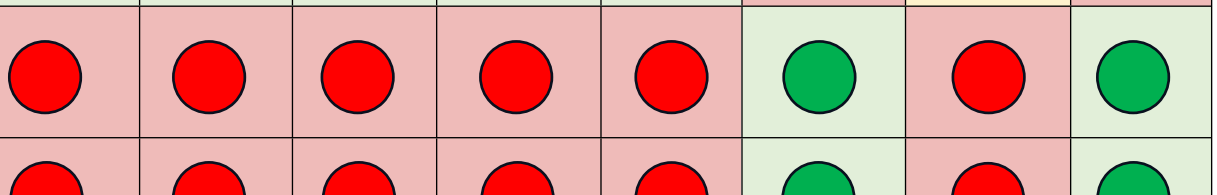
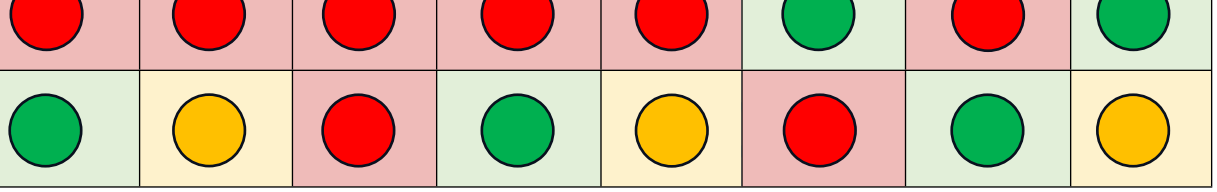
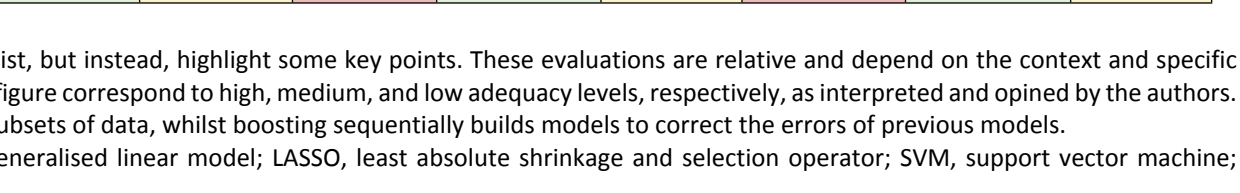


Figure 2. Overview of unsupervised ML scenario and commonly used algorithms [32–34]
Unsupervised learning clusters unlabeled data. The bottom panel provide an overview of the prevailing architectural concepts in ML employed to achieve specific objectives: A) K-means assigns instances to clusters based on feature similarity, using centroids for classification; B) Hierarchical clustering recursively generates nested sets of clusters in a dendrogram. Two strategies are employed: agglomerative clustering (bottom-up) and divisive clustering (top-down).

73x64mm (300 x 300 DPI)

AI and ML in febrile neutropenia

Table 1. A summary of standard ML models used in FN including intuition and key characteristics [2,32–34,44,45]

Model	Intuition	CHARACTERISTICS	Natural handling of data of "mixed" type	Handling of missing values	Robustness to outliers in input space	Computational scalability (large N)	Ability to deal with irrelevant inputs	Ability to extract linear combinations of features	Interpretability	Predictive power
Regression-based models [18,19,35–40]	These models assume a relationship between the response variable and a linear combination of predictors and can include penalty terms for regularisation and feature selection. Examples of regression models include logistic regression, penalised models (such as Ridge and LASSO) and GLM.									
Tree-based models [18,19,35–42]	These models are characterised by a hierarchical structure where internal nodes represent predictors and leaf nodes provide predictions. They can be effectively combined (ensemble) using techniques such as bagging (e.g., random forests) and boosting (e.g., GBM, XGBoost) *									
Artificial neural network (deep learning) [20,27,35,39,43]	A model inspired by the structure and function of the human brain, composed of interconnected layers of neurons									
SVM [18,35,37,39,41]	A model that employs an optimal boundary (or "hyperplane") to effectively classify data into distinct categories or classes by maximising the margin of separation between classes									
Naive Bayes [18,41]	A model that applies Bayes' theorem to calculate the probabilities of data points belonging to different classes and assign them to the class with the highest probability									

Note: The characteristics provided are not an exhaustive list, but instead, highlight some key points. These evaluations are relative and depend on the context and specific needs of the problem. Green, yellow, and red colors in the figure correspond to high, medium, and low adequacy levels, respectively, as interpreted and opined by the authors.
*Bagging combines multiple models trained on different subsets of data, whilst boosting sequentially builds models to correct the errors of previous models.
Abbreviations: GBM, gradient boosting machine; GLM, generalised linear model; LASSO, least absolute shrinkage and selection operator; SVM, support vector machine; XGBoost, extreme gradient boosting.

Table 2. Recently published ML approaches being undertaken to support clinical decision-making processes in FN

Reference	Main Goal	Applied Method	Nº	Results
Venäläinen *et al. [52]* *Cancer Med*, 2022	Predicting chemotherapy-induced FN	LASSO	Chemotherapy for non-haematological cancer patients: 5879. 2.24% experienced FN	Accurately predicted with an AUROC of 0.77
Cho *et al.* [35] *Sci Rep*, 2020	Predicting chemotherapy-induced FN	DT, XGBoost ANN, SVM, LASSO, Ridge	Breast cancer surgery in-patients receiving chemotherapy: 933. 409 (43.8%) experienced FN	Model performance based on AUROC values: DT (0.85), XGBoost (0.91), LASSO (0.86), SVM (0.88), ANN (0.86)
Wiberg *et al.* [36] *JCO Clin Cancer Inform*, 2021	Predicting chemotherapy-induced neutropenic events, including FN*	CART, OCT, LR/OFS**, RF, GBM	Haematological and non-haematological receiving chemotherapy: 2,806. Of the 17,513 encounters, 449 (2.6%) experienced FN	LR/OFS achieved an AUROC of 0.87 and an average precision of 0.15. The other models' AUROC values: RF (0.87), LR (0.86), XGB (0.82), OCT (0.80), CART (0.79)
Zhan *et al.* [41] Leuk Lymphoma, 2021	Predicting neutropenia and fever related to high-dose MTX	RF, SVM, NB, C5.0, rpart***	139 paediatric patients with newly diagnosed standard-/intermediate-risk B-cell ALL. 548 MTX treatment courses	RF was the top-performing model in predicting neutropenia with an AUROC of 0.93 and achieved high performance in predicting fever with an AUROCC of 0.87
Verma *et al.* [20] *IEEE J Transl Eng Health Med*, 2021	Wearable device for continuous temperature monitoring in high-risk FN	NARX****	86 patients post HSCT	Wearable device achieved a sensitivity of 90% and a specificity of 88% when detecting clinic-assessed fever episodes
Hwang *et al.* [27] *BMC Pulm Med*, 2021	Evaluate performance of CAD system in detecting pneumonia in FN	ANN-based CAD system	Total FN patients: 413. CXRs: 525	Average sensitivity of radiologists improved from 75% to 79% with the use of CAD. The CAD system achieved an AUROC of 0.89 and a specificity of 68%
Sung *et al. [37]* *BMC Cancer,* 2020	Detecting BSI in paediatric HSCT recipients	SVM, GBM, XGBoost, Elastic net*****	Patients included: 2,306. Eligible blood cultures: 11,183. Positive BSI: 624 (6%)	GBM achieved an AUROC of 0.74 and did not perform substantially better than using the presence of neutropenia alone to predict BSI
Alali *et al.* [38] *Sci Rep,* 2022	Detecting BSI and predicting admission to intensive care in paediatric patients with FN	RF, LR	Total FN episodes: 505 in 230 patients. Positive BSI: 106 (21%). Admission to intensive care: 56 (11%)	RF outperformed LR in predicting BSI and admission to intensive care, with AUROC of 0.79 vs 0.65 and 0.88 vs 0.76 respectively
Lind et al. [53] JAMA Netw Open, 2021	Assess high-risk bacteriemia associated sepsis and 10- and 28-day mortality in HSCT recipients	SuperLearner******	1943 HSCT recipients	SuperLearner achieved the best performance with an AUROC of 0.85 for high-risk bacteremia, 0.85 for 10-day mortality, and 0.80 for 28-day mortality

AI and ML in febrile neutropenia

Garcia-Vidal *et al.* [19] ***Infect Dis Ther*, 2021**	Assess risk of MDR-GNB infection at FN onset	RF, GBM, XGBoost, GLM	Total FN episodes: 3235 in 349 haematological patients. Data collected: ~7 million pieces of structured data	RF, GBM, XGBoost, and GLM achieved an AUROC of 0.79 with respective F1 scores of 0.97 for all models.
Padmanabhan *et al. [18]* ***Int. J. Environ. Res. Public Health*, 2022**	Assess risk of MDR, sepsis, and mortality in FN	GBM, XGBoost, SVM, LR, Ridge, NB	Total FN episodes: 1166 in 513 haematological patients	XGBoost achieved high predictive accuracy for clinical outcomes: sepsis (AUROC 0.85, recall 98%), MDR organisms (AUROC 0.91, recall 96%), and mortality (AUROC 0.88, recall 86%)
Du *et al.* [39] ***Int J Med Inform*, 2020**	Assess mortality risk in FN admissions without physician's subjective evaluation	Ridge, SVM, ANN, GBM	126,013 FN adult admissions	Ridge, SVM, ANN, and GBM achieved an AUROC of 0.92, with F1 scores ranging from 0.74 to 0.75
Hui *et al. [43]* ***Support Care Cancer*, 2011**	FN outcome prediction compared to MASCC risk index and Talcott model	ANN	Total FN patients: 227	ANN achieved an AUROC of 0.74 in predicting patients at low risk of complications or death, compared to MASCC (0.81) and Talcott (0.57)
Xiang *et al. [40]* ***Front Oncol*, 2021**	Assess septic shock risk in paediatric patients with fever or neutropenia and oncologic/haematological conditions	XGBoost, LR	1,238 children included Total septic shock: 64.	XGBoost achieved an AUROC of 0.93, outperforming pSOFA (0.76)
Jakob *et al.*† *[42]* ***Open Forum Infect Dis*, 2019**	Predicting death or admission to intensive care unit in prolonged FN	RF	211 (23%) adverse outcomes out of 927 episodes	RF achieved an AUC of 0.68

† Extracted from Open Forum Infectious Diseases, IDWeek 2019 Abstracts.

* The models were evaluated for the combined outcome of severe neutropenia and FN. No specific evaluation was conducted for FN alone. ** Selected model with 20 features using LR and OFS. *** Rpart and C5.0 are part of tree-based algorithms **** NARX (Nonlinear AutoRegressive with eXogenous input) is a recurrent ANN model used to capture nonlinear relationships in sequential data and time series. ***** Elastic net is a regularisation method that combines LASSO and Ridge techniques in regression models. ****** Ensemble learning algorithm in R that combines and weighs multiple ML models for enhanced predictive performance.

Abbreviations: ALL, acute lymphoblastic leukaemia; ANN, artificial neural network; AUROC, area under the receiver operating characteristic curve; BSI, bloodstream infection; CAD, computer-aided detection; CART, classification and regression trees; CXR, chest x-ray; DT, decision tree; FN, febrile neutropenia; GBM, gradient boosting machine; GLM, generalised linear model; HSCT, haematopoietic stem cell transplantation; LASSO, least absolute shrinkage and selection operator; LR, logistic regression; MASCC, Multinational Association for Supportive Care in Cancer; MDR-GNB, multidrug-resistant gram-negative bacteria; MTX, methotrexate; NB, naive bayes; OCT, optimal classification trees; OFS, optimal feature selection; pSOFA, paediatric sequential organ failure assessment; RF, random forest; rpart, recursive partitioning, regression trees; SVM, support vector machine; XGBoost, extreme gradient boosting.