An Explainable Machine Learning Model for Chronic Wound Management Decisions

Completed Research

Abstract

Recent advances in machine learning (ML) algorithms have motivated their use for automated Decision Support Systems (DSS). In healthcare domain, ML-based DSS enable providers to analyze large amounts of patient data and complex images quickly. However, providers find it difficult to interpret the predictions due to 'black box' ML reasonings. To facilitate meaningful interpretations, ML-based DSS should include explanation facilities as recommended in information systems (IS). For example, a wound care DSS should allow providers to see the reasoning based on wound location and presence of necrotic when it recommends referral. We present an ML-based DSS that provides global (reliance on the domain knowledge) and local (reasoning for predicting an instance) explanations for wound care decisions. We use Shapley values for a logistic regression (trained on wound visual features) which outperformed other classifiers in prediction of referral decisions (F-1 =0.938) and demonstrate its applicability in a wound care use-scenario. Finding suggest similar approach can be applied for other complex decision problems.

Keywords

Machine learning, Shapley explanations, interpretation, medical decision support, chronic wounds

Introduction

Clinical applications of artificial intelligence (AI) and ML techniques are receiving significant attention as medical professionals continue to investigate the accuracy of their diagnostic ability (Holzinger et al. 2019). In recent years, ML-based diagnostic systems have been developed to account for the errors in extremely ambiguous medical conditions (Latif et al. 2019) such as classifying age-related macular degeneration and diabetic macular edema (Kermany et al. 2018), diagnosing breast cancer (Wang et al. 2016), predicting the chance of developing pressure ulcers in hospitalized patients (Alderden et al. 2021) as well as patients at home (Orciuoli et al. 2020). Some of these diagnostic systems also assist medical professionals in their decision-making (Fatima et al. 2017). For instance, objective decisions about the severity level of Chronic Kidney Disease (CKD) were made possible based on predictions that use real-time data collected from patients' wearable IoT devices (Lakshmanaprabu et al. 2019). However, many of these ML methods are very complex and lack transparent (layman's) explanations for their results (Emmert-Streib et al. 2020).

The complexity usually comes from the high-dimensional input data on which these algorithms rely. As a result, even the cutting-edge ML DSS with high prediction power face adoption and usability challenges in real world medical decision-making tasks. The biggest failure is attributed to inability of such ML-based DSS in explaining their predictions in simple, context specific terms meaningful to less ML-savvy providers who constantly question the reasoning behind the ML predictions. In fact, recent regulations such as General Data Protection Regulation (in effect since 2018) force a mandatory inclusion of explanations as well as the right to challenge the decision outcome of their predictions (Kim et al. 2020).

In DSS literature, using explanations has been recommended since the advent of DSS (i.e., expert systems) to bring trust into the final prediction outcomes of DSSs (Kim et al. 2020; Nunes et al. 2017). The less intuitive and informative these prediction outcomes are, the less trustworthy they become for users who initially plan on adopting them. To make outcomes of a DSS intuitive, IS literature recommends supplementing the predictions outcome with one or more (1) definition explanations that supply descriptive or terminological information, (2) justifications explanations that rationalize part of the ML reasoning process by linking it to the global (domain) knowledge, (3) strategic explanations which explain the DSS overall problem-solving approach, and (4) rule-tracing explanations that explain why certain decision instances were or were not made by referring to the chunk of input data (Gregor and Benbasat 1999). For medical users who need to know why a certain patient has a certain (e.g., 30%) chance of developing an ulcer, justification and rule-tracing explanations are highly preferred (Darlington 2011; Holzinger et al. 2019). These explanations are commonly referred to as global and local explanations (Plumb et al. 2018).

While global explanations describe the model as a whole, in terms of which explanatory variables most determine its predictions, local explanations aim at interpreting individual predictions, at the single statistical unit level (Giudici and Raffinetti 2020). Since most real datasets have both continuous and discontinuous effects (due to mixture of categorical and numerical features), it is integral to design explanation systems that can capture, or are at least aware of, both types of effects (Plumb et al. 2018). This is typical for complex medical prediction tasks where data is noisy and highly non-linear (Shirwaikar et al. 2019). A ML-based wound DSSs are electronic systems designed to predict patient-specific assessments or recommendations by comparing patient wound characteristics with a wound knowledge base to directly assist wound care providers professionals in clinical decision such ad type of dressings, risk of ulceration, etc. at the time of decision-making and can be part of normal clinical workflow (Araujo et al. 2020). For example, in chronic wound decision making generating a decision for a wound requires consideration of one or more conditions (e.g., presence and amount of necrosis or unhealthy tissue and healthy granulation tissue, etc.) that may collectively influence the decision outcome. When a DSS is locally interpretable or equipped with rule-trace explanation, it allows providers to trace back the path of the predicted decision for a specific wound instance and to see which wound conditions were emphasized by the ML model during that instance (e.g., wound is located on plantar forefoot and is necrotic therefore it must be referred to wound clinic for debridement). Additionally, when DSS is complemented with a global explanation (that justifies model reliance on the domain knowledge or features), it can also display the overall wound conditions that are commonly known for triggering a sever wound needing specialist referral. Such explanations are needed to ensure that the prediction power of ML-based medical DSS is not underutilized.

We propose an explanation facility using Shapley values (from game theory that determines the contribution of features to the wound decision predictions by considering all possible combinations of healing features) for a wound care DSS tool that uses ML algorithms to predict care decisions based on visual wound features extracted from clinical images. The explanation facility provides global and local explanations for easier interpretation of decision predictions. We demonstrate how our ML-based DSS achieve high performance (F-1= .938) and high interpretability (Simply explainable wound care outcomes) through rigorous evaluations of our ML classifiers (i.e., cross validation, hyperparameter tuning and oversampling techniques) for a chronic wound decision-making scenario in rural underserved areas. Our main contribution is the application of state-of-the-art explainable solutions for a ML-based DSS assisting non-specialist providers in making complex chronic wound treatment decisions. These explanations are designed to help non-specialist providers avoid errors when interpreting the decision predictions. Our approach will guide future IS researchers in developing similar explainable methods for their predictive DSS solutions.

Background

Use-scenario

Chronic wounds or ulcers affect 6.5 million Americans (Fife 2012) and can be extremely difficult to treat (Han 2017) with costs of $32 billion annually (Nussbaum 2018). Untreated or inappropriately treated chronic wounds (Kim 2013; Wu 2010) can endanger patients' lives (Järbrink 2017) or result in undesired healing outcomes with high risk of amputation (Jeffcoate 2004) following limited quality of life. Thus, patients must receive appropriate treatment to maintain normal healing process (Frykberg 2015). Unfortunately, chronic wound patients in remote underserved areas do not always have access to evidence-based wound care services or specialty wound clinics (Benskin 2013). These patients usually receive their routine care from non-specialist providers who have limited chronic wound assessment experience (McIntosh 2008). Even though published chronic wound management guides are abundant, non-specialist providers face major decision uncertainties when it comes to a particular wound (Christie 2018).

We developed a ML-based DSS solution to support these providers. Our smartphone app equipped a reasoning engine, allows the non-specialist provider (visiting a patient in rural underserved area) to capture a photo of the patient's wound after the initial examination. Then, the wound photo is processed using image segmentation algorithms for tissue analysis and prediction of wound severity score (PWAT) that is validated for use in photo-based wound assessment (Thompson et al. 2013). Next, the non-specialist provider selects the presence of basic wound features that impede healing (i.e., ischemia, rolled edge, etc.) as well as the wound location. Finally, the ML-based prediction model predicts a decision outcome as whether to refer the wound patient to a wound specialist or advise on continuing with current treatment plan (once, PWAT, location and impeding features are inputted and a decision on Referral or Non-referral is generated). These decisions will be predicted based on patient's wound conditions (e.g., necrotic, or unhealthy tissue amount, epithelium & granulation healthy tissues, wound location, etc.) during each visit. Furthermore, the DSS will provide explanations that allow the visiting nurse to locally track and interpret patient's specific wound conditions and see how much model relied on global domain specific features (Wound bed, edge, and skin features) at the time of care.

The need for explanations

Explanations in IS are intended for any recommendation system that is designed to provide users decision support in any given decision context. ML explanations can be based on transparency of input data or the ML model itself. Data transparency ensures that training process for classification was done properly and poor results, if any, are not due to poorly labeled data (He 2019). Model transparency, on the other hand, ensures that a certain decision path predicted by medical ML-based DSS is interpretable, logical and trustworthy (He 2019), and conforms with ethical requirements and medical regulations (Ahmad 2018). Hence, it is important for a ML-based medical DSS to provide explanations based on model transparency using both local and global techniques. The global techniques have the advantage of being generalizable to the entire population while local interpretability techniques focus on giving explanations at the level of instances (Elshawi et al. 2019). Both methods can be equally valid and effective for assisting providers medical decision process. However, the providers will always remain to hold the final say on accepting or rejecting the outcome of the ML-based DSS and its explanations using their domain expertise (Elshawi et al. 2019).

Prior work

Several AI and ML-based studies explored the use of explanation techniques for medical diagnostic tasks. One study (Schulz et al. 2019) trained a diagnostic classifier (healthy vs. diseased) and extracted instance-wise explanations for the classifier's decisions on identification of disease subtypes and corresponding biomarkers that can substantially improve clinical diagnosis and treatment selection. The authors used game-theory Shapley values (determining the relative contribution of variables to the predictions by considering all possible combinations of variables that maximize payoff) to generate instance-wise explanations for the predictions of the classifier. In another study Athanasiou et al. (2020 developed and evaluated an explainable risk support model for predicting the fatal or non-fatal Cardiovascular Disease (CVD) incidence in individuals with Type 2 diabetes. Their approach was based on the XGBoost machine learning classification algorithm and Shapley explanations (Zihni et al. 2020) that support calculation of the 5-year CVD risk and the generation of instance-based explanations on the model's decisions. Although these studies demonstrated the validity of their explanations approaches (visually projecting the most important features or components involved in predictions), they targeted users already knowledgeable in Machine Learning. Specifically, they failed to provide actionable (directly usable) explanations that facilitate the understanding, and thus objective decision making, of non-specialist providers who are not familiar with ML jargon. Additionally, these type of explanations generally describe "why" a prediction is made but lack in explaining "how" the decision is made (Mohseni et al. 2018). In the context of chronic wound management, we have not found any study that used explanations for users especially non-specialist' providers who are typically have limited knowledge of ML. Majority of the proposed ML-based methods for chronic wound management focused on wound tissue (healthy granulation and unhealthy necrosis) segmentation (Chakraborty 2016), skin tear classification (Nagata et al. 2021), and wound size measurement using smartphone camera (Lucas et al. 2020).

To address the above-mentioned gaps, and in response to recent calls encouraging explainable ML for healthcare (Ahmad 2018; Holzinger et al. 2020; Mohseni et al. 2018), our ML-based wound care DSS solution generates transparent explanations (He 2019; Holzinger 2018; Holzinger 2017) that show how each wound care decision is predicted. As mentioned earlier, extent wound care decision guidelines exist for bedside wound care (Franks 2016; Kottner 2019) with narrative explanations about assessment procedures and generalized recommendations about which decision to take. However, their use is too limited to be considered for actionable and objective wound care decisions. Hence, alternative solutions using smartphone and ML algorithms can enhance wound healing, reduce costs (Nasi 2015) and enhance simple interpretation of predicted decisions for non-specialists users.

Methodology

Data collection and preparation

Figure 1 presents our approach to training our explainable ML classifiers. First, we collected 2056 raw images and used the only validated and photo-based wound scoring tool called PWAT (Thompson et al. 2013) and its image criteria to exclude bad quality images and finally included 375 wound images for training. PWAT scores eight attributes of wounds, (1) size, (2) depth, (3,4) type and amount of necrotic tissue, (5,6) type and amount of granulation tissue, (7) wound edges and (8) peri-ulcer skin viability (Thompson et al. 2013). Our local wound experts (plastic surgeon and nurse practitioner) labeled these raw wound images with referral and non-referral decisions. Additionally, we created evidence-based labels using standard wound care decision-making guidelines. Finally, we applied majority voting rule across all the given labels to create our final decision labels. Since the focus of this paper is the training and evaluation of explainable ML classifiers, we will refer the reader to our prior publications where we provide a detailed description of our approach for data collection, preparation, and labeling as well as the performance of our predictions (BlindReview).

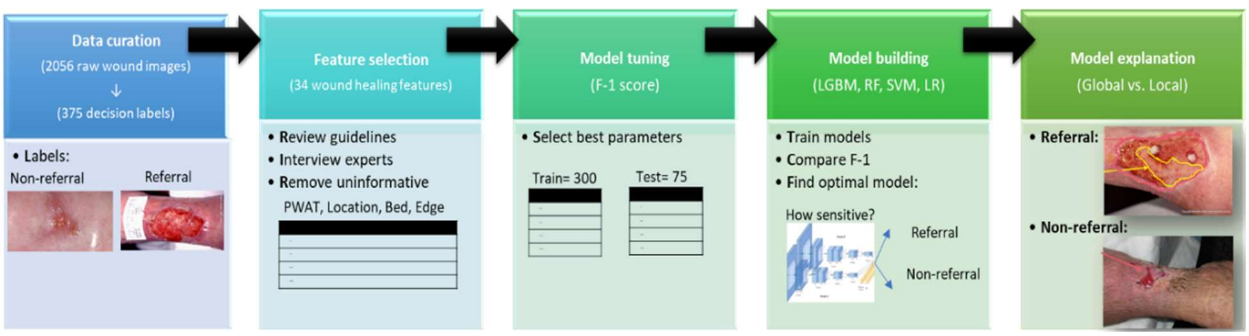

Figure 1. Interpretable ML decision support development process

Feature selection

We included wound severity measure PWAT which scores a wound from *0* (or healing) to *32* (or extremely bad) conditions. We also searched for additional wound features within wound care guidelines and after removing uninformative features (using mutual information and Pearson correlation) we prepared our final dataset including PWAT total score, wound location, bed, and edge features. Our decision classes are "Referral" for when a wound is severely damaged or "Non-referral" for when the wound patient may continue with current plan due to healing status. These decisions were derived from our original decisions based on standard of wound care (i.e., continue with current treatment, request change from a wound specialist or refer patient to a wound specialist). This was done since for medical classification problems the correct diagnosis in a particular case and what constitutes best practice can be controversial (Char 2018). Otherwise, training ML classifiers can be challenging due to difficulty of data integration (Holzinger 2018), the fact that wounds corresponding to different decisions can appear quite similar (fuzzy decision boundaries), and the multi-label nature of decision classes (i.e. more than one correct possible decision path exists for a wound. For example, for a seemingly large open wound a plastic surgeon may suggest skin grafting while a vascular surgeon may recommend vacuum closure. We expected this to be an issue thus experimented with binary classification for more objective, interpretable decisions.

ML classifiers

Extant literature exists on types of ML classifiers suitable for supporting medical decisions where predictions are costly. To build our DSS we used Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR) and *Light Gradient Boosting Model (LGBM)* as a novel gradient boosting decision tree algorithm (Ke 2017; Sun 2020). Prior classification studies achieved high prediction accuracy (i.e., above 80%) when using these classifiers (Portugal et al. 2018). For tuning the hyperparameters of these classifiers we used Bayesian Optimization search from open-source Ray tune-Sklearn python package (Liaw et al. 2018) that find the best performing parameters in as few iterations as possible (Perrone et al. 2019) based on Bayesian inference and Gaussian processes. The hyperparameter tuning was performed using a 5-fold cross validation on 300 data samples for training and 75 for testing.

Performance metrics

Since we deal with cost-sensitive classification problem, the first and the most preliminary issue is how to measure the errors in such a way that no superficial, often wrong, constraints or assumptions are imposed (Ben-David 2008). We first balanced our dataset using Synthetic Minority Over-sampling Technique (SMOTE) (Chawla 2002) which created new artificial wound data samples (Buda 2018). This was done together with 5-fold cross validation scheme. We then chose Weighted F-1 score (weighted by the number of true instances for each label) as our main metric of ML classifiers comparison since it is highly recommended for our imbalanced classes.

$$F - 1 = 2\frac{Precision\ .\ Recall}{Precision + Recall}$$

Additionally, we use Area Under the Receiver Operating Characteristics Curve (AUC) which is a widely used measure of performance for supervised classification problems, mainly applicable to binary classifications (Hand 2001). When AUC= 0 the ML-based DSS has no screening ability, if AUC= 0.5 the DSS is as good as random guessing, and if AUC = 1, DSS can perfectly distinguish “Referral” and “Non-referral” wound conditions. We also included sensitivity measure (true positive rates) to identify the classifiers with high power of screening referral patients to validate and recommended their use for our wound decision prediction.

Explanation method

Several explanation techniques such as LIME, MAPLE, Anchors and Shapley have been proposed in the literature to facilitate the interpretation of ML models (ElShawi et al. 2020). However, recent literature recommends the use of Shapley (based on mean Shapley values) since it provides both local and global interpretation as well as better detection of ML bias (ElShawi et al. 2020). When applied to wound decision predictions, a non-specialist provider will be able to identify incorrect predictions by comparing to global wound healing features recommended in the guidelines. Hence, to comply with latest literature for medical diagnosis prediction (Ahmad 2018) our ML-based DSS is designed to explain each predicted wound decision both globally and locally. To provide global explanations, we use Shapley feature importance that indicate which wound conditions have the highest global impacts on each decision path. Shapley feature importance, takes a game theory approach to determine the relative contribution of variables to the predictions by considering all possible combinations of variables as cooperating and competing coalitions to maximize payoff, defined in terms of the prediction (Ahmad 2018) based on a formula:

$$\emptyset_i = \sum_{S \subset \setminus F\{i\}.} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f_{S\cup\{i\}}\,(x_{S\cup\{i\}}) - f_S\,(x_S)\right].$$

where |F| is the size of the full coalition (power of the complete feature set), S is any subset of the coalition that does not include feature i, and |S| is the size of that subset. The bit (·) at the end is how much larger is the payoff when we add player i to this subset S. The Shapley approach enables the identification and prioritization of features (i.e., wound healing features that positively or negatively affected the ML-based DSS decision for the nurse to see in our use case) and can be applied to any ML model (Rodríguez-Pérez and Bajorath 2020). Additionally, we compare and validate our Shapley feature importance (global) with permutation importance as a means of confirming wound domain specific wound features that were triggered by the model.

Results

Table 1 presents classifiers' performance results for wound prediction along with tuned hyper-parameters. According to this table, all classifiers outperformed the baseline random classifier with a large gap. The overall highest performance was achieved by LR with F-1 score of 0.938 and AUC of 0.896 in predicting wound decisions on "referral" or "non-referral" cases. LR is recommended for screening referral wound care decisions due to its high sensitivity. These decisions are critical for non-specialist nurses dealing with wound patients who cannot afford regular visits to specialist clinic unless their wound is in a critical condition that requires an urgent referral.

ML Model (The hyper-parameters were tuned based on F-1 as scoring metric)	**Sensitivity** **(Referral)**	**Sensitivity** **(Non-referral)**	**F-1**	**AUC**
LR (C = '100', penalty solver= 'l2', solver= 'liblinear')	**.963**	**.828**	**.938**	**.896**
SVM (kernel='rbf', C='361', gamma='0.005)	.937	.828	.917	.883
LGBM (learning_rate= '1', n_estimators= '260', num_leaves= '247', boosting_type= 'gbdt')	.940	.728	.901	.834
RF (n_estimators= '95', criterion= 'gini', max_depth= '18', min_samples_leaf = '5', min_samples_split= '8')	.908	.785	.904	.852
Dummy classifier (Baseline)	.501	.542	.509	.522

Table 1. ML classifier performance comparison.

We were able to demonstrate explainability of LR by using Shapley and permutation feature importance for global (*LR reliance on wound healing features*) and Shapley for local (*LR reasoning for predicting a referral case*) explanations (Figure 2). Shapley importance (mean values across both decision classes) as shown in figure 2-a demonstrates that LR relied mainly on wound bed followed by wound edge and PWAT severity score. This was expected since these wound features are known to indicate healing within standard clinical guidelines. Next, we used permutation importance to confirm the results of Shapley feature importance. As shown in figure 2-b, the same three wound healing features were identified as top features that LR relied on for prediction even after permutation, that is, shuffling the features in each model training iteration. However, both approaches confirmed that the bony prominence feature was less informative in assisting LR to predict referral or non-referral cases. This is in line with standard wound care guidelines where the boney prominence feature is used interchangeably with any wound location spotted on a bony part of the foot. For example, a wound located on the ankle is, by definition, a bony prominence wound. We will consider this feature combination in our next iteration of the wound care prediction systems. Additionally, we illustrated actual Shapley values (Figure 2-c) across encoded features which indicate positive and negative impact on the "Referral" decision predictions along with their high or low magnitude in red and blue, respectively.

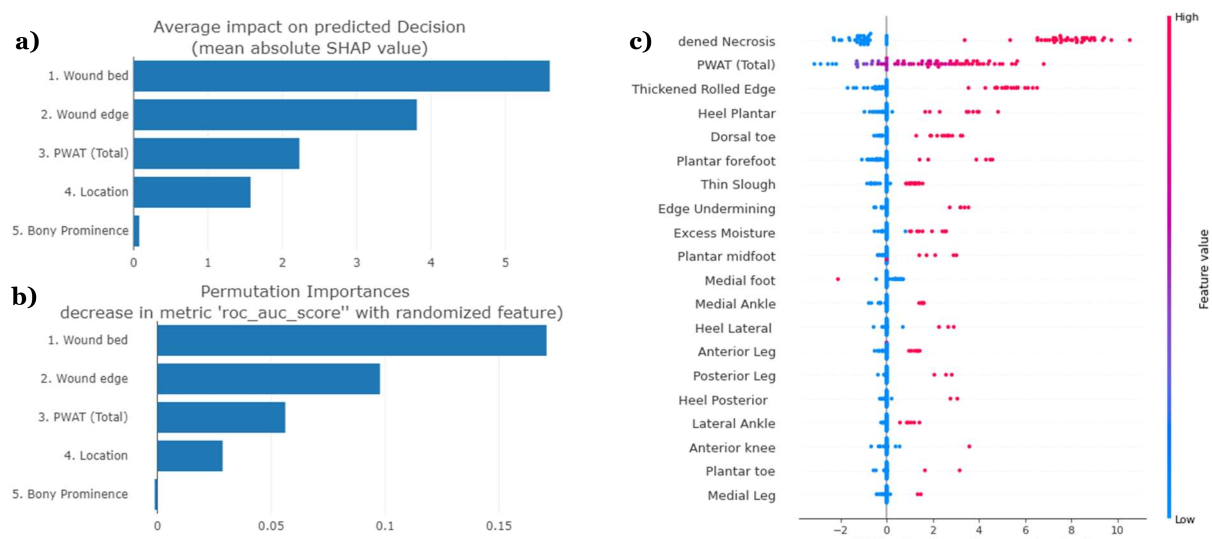

Figure 2. Shapley importance for wound care confirmed by permutation importance

A ML-based wound DSS may be equipped with local explanation facility illustrating reasoning behind a decision recommendation for a wound instance with information about model reliance on certain wound domain features (e.g., wound healing features) for the nurse to consider. In our use-scenario, LR-based DSS can guide the visiting nurse both in choosing the right decision path and interpreting it. For instance, during a patient home visit a nurse may discover a wound located on the lateral ankle (Figure 3-a).

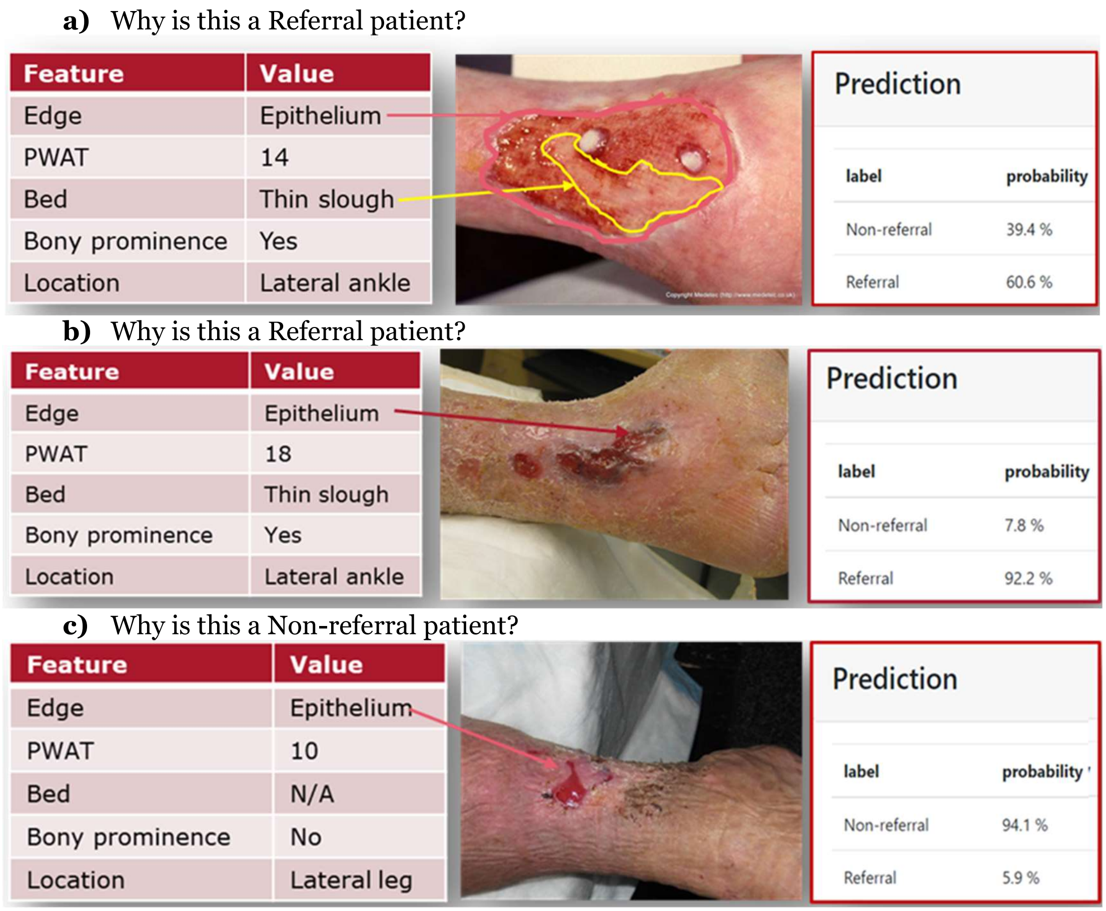

Figure 3. Shapley explanations for a wound instance

After initial assessment, the nurse determines that the wound looks healthy due to presence of epithelium on the wound edge. When the LR-based DSS conflicts the non-specialist nurse's assessment by predicting referral with approximately 61% chance (due to high reliability of the LR predictions on wound healing features), the explanation facility of the DSS becomes critical. For instance, in our use-case, explanations demonstrate that although the wound edge is healthy, the wound is severe enough (PWAT = 14) to be considered for a referral (i.e., to be debrided or cleaned) due to presence of thin slough which impedes wound healing. Given 61% chance, the nurse may decide to contact the wound specialist before referring the patient. However, if the chance of referral presented in the explanation facility was higher (Figure 3-b), the nurse may refer the patient to a specialist. Similarly, for a wound located on the lateral leg (Figure 3-c), LR-based DSS recommends a firm "Non-referral" (i.e., there is roughly 6% chance of referral) due to low severity of the wound (PWAT=10) and healthy wound bed (i.e., no presence of unhealthy tissues). Hence, the nurse can collect these wound measurements and advise the patient on continuing with the current treatment.

Conclusion and limitation

ML-based DSS for medical decision-making usually fail to provide simple interpretations for their predicted outcomes. As a result, most of these medical DSS are considered less reliable and usable for non-specialist providers who seek simplicity in DSS outcomes due to lack of expertise in one specific medical domain. To address this problem, we designed an explainable ML solution usable for complex medical decision-making problems. We used both global and local explanation solutions to demonstrate how our ML-based DSS can be both high-performing and interpretable. We applied these solutions for chronic wound management decisions as a use case. Findings suggest visiting non-specialist nurses can simply interpret and hence generalize the predictions for any given wound decision scenario. We provided clear instances illustrating how these predictions and their explanations can be interpreted if designed and integrated into an envisioned smartphone DSS App supporting non-specialists in their chronic wound decision making. A similar approach can be used for the development of ML-based DSS tools assisting novice decision makers across those disciplines that use a high-level decision process and multi-factor complex decision-making criteria.

This research has some limitations. We were limited to a training dataset of 375 wound images which may be considered small and a threat to generalizability of the ML-based DSS envisioned for a real-world setting. To avoid this issue, we oversampled the data and ran a thorough evaluation that demonstrated our classifiers performance using cross-validation. However, future research can further verify our classifiers using larger datasets. Collecting, labeling, and validating wound care datasets is challenging due to the limited availability of wound image datasets with wound care decisions. Additionally, manual feature extraction from wound images is not considered highly efficient for clinical use of our DSS. Thus, we are experimenting with automatic feature extractors using deep learning to be integrated as part of feature selection for our ML-based DSS. Our future work includes integration of these ML techniques into a smartphone App using image processing techniques that allow for automatic feature extraction of our predictors for these decision-making policies. We also plan on designing and integrating an explanation interface to be integrated in the DSS App.

REFERENCES

Ahmad, M. A. E., Carly: Teredesai, Ankur. 2018. "Interpretable Machine Learning in Healthcare," *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*: ACM, pp. 559-560.

Alderden, J., Drake, K. P., Wilson, A., Dimas, J., Cummins, M. R., Yap, T. L. J. B. M. I., and Making, D. 2021. "Hospital Acquired Pressure Injury Prediction in Surgical Critical Care Patients," (21:1), pp. 1-11.

Araujo, S. M., Sousa, P., and Dutra, I. J. J. m. i. 2020. "Clinical Decision Support Systems for Pressure Ulcer Management: Systematic Review," (8:10), p. e21621.

Athanasiou, M., Sfrintzeri, K., Zarkogianni, K., Thanopoulou, A. C., and Nikita, K. S. 2020. "An Explainable Xgboost–Based Approach Towards Assessing the Risk of Cardiovascular Disease in Patients with Type 2 Diabetes Mellitus," *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*: IEEE, pp. 859-864.

Ben-David, A. 2008. "Comparison of Classification Accuracy Using Cohen's Weighted Kappa," *Expert Systems with Applications* (34:2), pp. 825-832.

Benskin, L. 2013. "A Review of the Literature Informing Affordable, Available Wound Management Choices for Rural Areas of Tropical Developing Countries," *Ostomy/wound management* (59:10), pp. 20-41.

Buda, M. M., Atsuto: Mazurowski, Maciej A. 2018. "A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks," *Neural Networks* (106), pp. 249-259.

Chakraborty, C. G., Bharat: Ghosh, Soumya K: Das, Dev K: Chakraborty, Chandan. 2016. "Telemedicine Supported Chronic Wound Tissue Prediction Using Classification Approaches," *Journal of medical systems* (40:3), p. 68.

Char, D. S. S., Nigam H: Magnus, David. 2018. "Implementing Machine Learning in Health Care—Addressing Ethical Challenges," *The New England journal of medicine* (378:11), p. 981.

Chawla, N. V. B., Kevin W: Hall, Lawrence O: Kegelmeyer, W Philip. 2002. "Smote: Synthetic Minority over-Sampling Technique," *Journal of artificial intelligence research* (16), pp. 321-357.

Christie, J. G., Trish A: Dumville, Jo C: Cullum, Nicky A. 2018. "Do Systematic Reviews Address Community Healthcare Professionals' Wound Care Uncertainties? Results from Evidence Mapping in Wound Care," *PloS one* (13:1), p. e0190045.

Darlington, K. W. J. S. O. 2011. "Designing for Explanation in Health Care Applications of Expert Systems," (1:1), p. 2158244011408618.

Elshawi, R., Al-Mallah, M. H., Sakr, S. J. B. m. i., and making, d. 2019. "On the Interpretability of Machine Learning-Based Model for Predicting Hypertension," (19:1), pp. 1-32.
ElShawi, R., Sherif, Y., Al-Mallah, M., and Sakr, S. J. C. I. 2020. "Interpretability in Healthcare: A Comparative Study of Local Machine Learning Interpretability Techniques,").
Emmert-Streib, F., Yli-Harja, O., Dehmer, M. J. W. I. R. D. M., and Discovery, K. 2020. "Explainable Artificial Intelligence and Machine Learning: A Reality Rooted Perspective," (10:6), p. e1368.
Fatima, M., Pasha, M. J. J. o. I. L. S., and Applications. 2017. "Survey of Machine Learning Algorithms for Disease Diagnostic," (9:01), p. 1.
Fife, C. E. C., Marissa J: Walker, David: Thomson, Brett. 2012. "Wound Care Outcomes and Associated Cost among Patients Treated in Us Outpatient Wound Centers: Data from the Us Wound Registry," *Wounds* (24:1), p. 10.
Franks, P. J. B., Judith: Collier, Mark: Gethin, Georgina: Haesler, Emily: Jawien, Arkadiusz: Laeuchli, Severin: Mosti, Giovanni: Probst, Sebastian: Weller, Carolina. 2016. "Management of Patients with Venous Leg Ulcers: Challenges and Current Best Practice," *Journal of wound care* (25:Sup6), pp. S1-S67.
Frykberg, R. G. B., J. 2015. "Challenges in the Treatment of Chronic Wounds," *Adv Wound Care (New Rochelle)* (4:9), pp. 560-582.
Giudici, P., and Raffinetti, E. J. E. S. w. A. 2020. "Shapley-Lorenz Explainable Artificial Intelligence,"), p. 114104.
Gregor, S., and Benbasat, I. J. M. q. 1999. "Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice,"), pp. 497-530.
Han, G. C., R. 2017. "Chronic Wound Healing: A Review of Current Management and Treatments," *Adv Ther* (34:3), pp. 599-610.
Hand, D. J. T., Robert J %J Machine learning. 2001. "A Simple Generalisation of the Area under the Roc Curve for Multiple Class Classification Problems," (45:2), pp. 171-186.
He, J. B., Sally L: Xu, Jie: Xu, Jiming: Zhou, Xingtao: Zhang, Kang. 2019. "The Practical Implementation of Artificial Intelligence Technologies in Medicine," *Nature medicine* (25:1), p. 30.
Holzinger, A. 2018. "From Machine Learning to Explainable Ai," *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*: IEEE, pp. 55-66.
Holzinger, A., Carrington, A., and Müller, H. J. K.-K. I. 2020. "Measuring the Quality of Explanations: The System Causability Scale (Scs),"), pp. 1-6.
Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H. J. W. I. R. D. M., and Discovery, K. 2019. "Causability and Explainability of Artificial Intelligence in Medicine," (9:4), p. e1312.
Holzinger, A. B., Chris: Pattichis, Constantinos S: Kell, Douglas B. 2017. "What Do We Need to Build Explainable Ai Systems for the Medical Domain?," *arXiv preprint arXiv:1712.09923*).
Järbrink, K. N., Gao: Sönnergren, Henrik: Schmidtchen, Artur: Pang, Caroline: Bajpai, Ram: Car, Josip. 2017. "The Humanistic and Economic Burden of Chronic Wounds: A Protocol for a Systematic Review," *Systematic reviews* (6:1), p. 15.
Jeffcoate, W. J. v. H., W. H. 2004. "Amputation as a Marker of the Quality of Foot Care in Diabetes," *Diabetologia* (47:12), pp. 2051-2058.
Ke, G. M., Qi: Finley, Thomas: Wang, Taifeng: Chen, Wei: Ma, Weidong: Ye, Qiwei: Liu, Tie-Yan. 2017. "Lightgbm: A Highly Efficient Gradient Boosting Decision Tree," *Advances in neural information processing systems*, pp. 3146-3154.
Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., and Yan, F. J. C. 2018. "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," (172:5), pp. 1122-1131. e1129.
Kim, B., Park, J., and Suh, J. J. D. S. S. 2020. "Transparency and Accountability in Ai Decision Support: Explaining and Visualizing Convolutional Neural Networks for Text Information," (134), p. 113302.
Kim, P. J. E., K. K.: Steinberg, J. S.: Pollard, M. E.: Attinger, C. E. 2013. "Critical Elements to Building an Effective Wound Care Center," *J Vasc Surg* (57:6), pp. 1703-1709.
Kottner, J. C., Janet: Carville, Keryln: Balzer, Katrin: Berlowitz, Dan: Law, Susan: Litchford, Mary: Mitchell, Pamela: Moore, Zena: Pittman, Joyce. 2019. "Prevention and Treatment of Pressure Ulcers/Injuries: The Protocol for the Second Update of the International Clinical Practice Guideline 2019," *Journal of tissue viability* (28:2), pp. 51-58.
Lakshmanaprabu, S., Mohanty, S. N., Krishnamoorthy, S., Uthayakumar, J., and Shankar, K. J. A. S. C. 2019. "Online Clinical Decision Support System Using Optimal Deep Neural Networks," (81), p. 105487.

Latif, J., Xiao, C., Imran, A., and Tu, S. 2019. "Medical Imaging Using Machine Learning and Deep Learning Algorithms: A Review," *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*: IEEE, pp. 1-5.

Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J. E., and Stoica, I. J. a. p. a. 2018. "Tune: A Research Platform for Distributed Model Selection and Training,").

Lucas, Y., Niri, R., Treuillet, S., Douzi, H., and Castaneda, B. J. A. i. w. c. 2020. "Wound Size Imaging: Ready for Smart Assessment and Monitoring,").

McIntosh, C. O., Karen. 2008. "A Survey of Nurses' and Podiatrists' Attitudes, Skills and Knowledge of Lower Extremity Wound Care," *Wounds UK* (4:1).

Mohseni, S., Zarei, N., and Ragan, E. D. J. a. p. a. 2018. "A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable Ai Systems,").

Nagata, T., Noyori, S. S., Noguchi, H., Nakagami, G., Kitamura, A., and Sanada, H. J. J. o. T. V. 2021. "Skin Tear Classification Using Machine Learning from Digital Rgb Image,").

Nasi, G. C., Maria: Guerrazzi, Claudia. 2015. "The Role of Mobile Technologies in Health Care Processes: The Case of Cancer Supportive Care," *Journal of medical Internet research* (17:2), p. e26.

Nunes, I., Jannach, D. J. U. M., and Interaction, U.-A. 2017. "A Systematic Review and Taxonomy of Explanations in Decision Support and Recommender Systems," (27:3), pp. 393-444.

Nussbaum, S. R. C., M. J.: Fife, C. E.: DaVanzo, J.: Haught, R.: Nusgart, M.: Cartwright, D. 2018. "An Economic Evaluation of the Impact, Cost, and Medicare Policy Implications of Chronic Nonhealing Wounds," *Value Health* (21:1), pp. 27-32.

Orciuoli, F., Orciuoli, F. J., and Peduto, A. J. P. C. S. 2020. "A Mobile Clinical Dss Based on Augmented Reality and Deep Learning for the Home Cares of Patients Afflicted by Bedsores," (175), pp. 181-188.

Perrone, V., Shcherbatyi, I., Jenatton, R., Archambeau, C., and Seeger, M. J. a. p. a. 2019. "Constrained Bayesian Optimization with Max-Value Entropy Search,").

Plumb, G., Molitor, D., and Talwalkar, A. J. a. p. a. 2018. "Model Agnostic Supervised Local Explanations,").

Portugal, I., Alencar, P., and Cowan, D. J. E. S. w. A. 2018. "The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review," (97), pp. 205-227.

Rodríguez-Pérez, R., and Bajorath, J. J. J. o. c.-a. m. d. 2020. "Interpretation of Machine Learning Models Using Shapley Values: Application to Compound Potency and Multi-Target Activity Predictions," (34:10), pp. 1013-1026.

Schulz, M.-A., Chapman-Rounds, M., Verma, M., Bzdok, D., and Georgatzis, K. J. a. p. a. 2019. "Clusters in Explanation Space: Inferring Disease Subtypes from Model Explanations,").

Shirwaikar, R. D., Acharya, D., Makkithaya, K., Surulivelrajan, M., and Srivastava, S. J. A. I. i. m. 2019. "Optimizing Neural Networks for Medical Data Sets: A Case Study on Neonatal Apnea Prediction," (98), pp. 59-76.

Sun, X. L., Mingxi: Sima, Zeqian %J Finance Research Letters. 2020. "A Novel Cryptocurrency Price Trend Forecasting Model Based on Lightgbm," (32), p. 101084.

Thompson, N., Gordey, L., Bowles, H., Parslow, N., Houghton, P. J. A. i. s., and care, w. 2013. "Reliability and Validity of the Revised Photographic Wound Assessment Tool on Digital Images Taken of Various Types of Chronic Wounds," (26:8), pp. 360-373.

Wang, D., Khosla, A., Gargeya, R., Irshad, H., and Beck, A. H. J. a. p. a. 2016. "Deep Learning for Identifying Metastatic Breast Cancer,").

Wu, S. C. M., W.: Armstrong, D. G. 2010. "Wound Care: The Role of Advanced Wound Healing Technologies," *J Vasc Surg* (52:3 Suppl), pp. 59S-66S.

Zihni, E., Madai, V. I., Livne, M., Galinovic, I., Khalil, A. A., Fiebach, J. B., and Frey, D. J. P. o. 2020. "Opening the Black Box of Artificial Intelligence for Clinical Decision Support: A Study Predicting Stroke Outcome," (15:4), p. e0231166.