World Academy of Science, Engineering and Technology
International Journal of Sport and Health Sciences
Vol:19, No:10, 2025

Using Machine Learning to Extract Patient Data from Non-standardized Sports Medicine Physician Notes

Authors : Thomas Q. Pan, Anika Basu, Chamith S. Rajapakse

Abstract : Machine learning requires data that is categorized into features that models train on. This topic is important to the field of sports medicine due to the many tools it provides to physicians such as diagnosis support and risk assessment. Physician note that healthcare professionals take are usually unclean and not suitable for model training. The objective of this study was to develop and evaluate an advanced approach for extracting key features from sports medicine data without the need for extensive model training or data labeling. An LLM (Large Language Model) was given a narrative (Physician's Notes) and prompted to extract four features (details about the patient). The narrative was found in a datasheet that contained six columns: Case Number, Validation Age, Validation Gender, Validation Diagnosis, Validation Body Part, and Narrative. The validation columns represent the accurate responses that the LLM attempts to output. With the given narrative, the LLM would output its response and extract the age, gender, diagnosis, and injured body part with each category taking up one line. The output would then be cleaned, matched, and added to new columns containing the extracted responses. Five ways of checking the accuracy were used: unclear count, substring comparison, LLM comparison, LLM re-check, and hand-evaluation. The unclear count essentially represented the extractions the LLM missed. This can be also understood as the recall score ([total - false negatives] over total). The rest of these correspond to the precision score ([total - false positives] over total). Substring comparison evaluated the validation (X) and extracted (Y) columns' likeness by checking if X's results were a substring of Y's findings and vice versa. LLM comparison directly asked an LLM if the X and Y's results were similar. LLM Re-check prompted the LLM to see if the extracted results can be found in the narrative. Lastly, A selection of 1,000 random narratives was also selected and hand-evaluated to give an estimate of how well the LLM-based feature extraction model performed. With a selection of 10,000 narratives, the LLM-based approach had a recall score of roughly 98%. However, the precision scores of the substring comparison and LLM comparison models were around 72% and 76% respectively. The reason for these low figures is due to the minute differences between answers. For example, the 'chest' is a part of the 'upper trunk' however, these models cannot detect that. On the other hand, the LLM re-check and subset of hand-tested narratives showed a precision score of 96% and 95%. If this subset is used to extrapolate the possible outcome of the whole 10,000 narratives, the LLM-based approach would be strong in both precision and recall. These results indicated that an LLM-based feature extraction model could be a useful way for medical data in sports to be collected and analyzed by machine learning models. Wide use of this method could potentially increase the availability of data thus improving machine learning algorithms and supporting doctors with more enhanced tools.

Keywords : AI, LLM, ML, sports
Conference Title : ICAIS 2025 : International Conference on Artificial Intelligence in Sports
Conference Location : Tunis, Tunisia
Conference Dates : October 25-26, 2025

Open Science Index, Sport and Health Sciences Vol:19, No:10, 2025 publications.waset.org/abstracts/195463.pdf