INTRODUCTION

Enh**anced** **R**ule**- and** **L**ogic**-based** **I**mmunological **C**onsultative **H**ub or **EHRLICH**, a reverse acronym inspired by the German pioneer of immunology Paul Ehrlich, is an expert system specializing in the following diseases: Addison disease, dermatomyositis, Hashimoto thyroiditis, multiple sclerosis, Crohn's disease, Grave's disease, myasthenia gravis, pernicious anemia, reactive arthritis, Sjögren syndrome, lupus, rheumatoid arthritis, type I diabetes, celiac disease, and alopecia areata.

The specific domain of the system is **autoimmune diseases**, which are triggered by malfunctions of the immune system that cause it to attack healthy body tissues. In particular, its goal is to automate the diagnosis for 15 such diseases, selected on the basis of their prevalence and the accessibility of symptomatic detection (i.e., symptoms that may be indicative of their presence can be determined to an extent with visual inspection or the patient's own assessment prior to conclusive laboratory tests).

Autoimmune diseases are some of the most commonly misdiagnosed illnesses, highlighting the need for expertise in this domain, and immunologists themselves require roughly 12 years of education and training before recognition (Fischer, 2011). Given the prevalence of undiagnosed diseases in the Philippines, with more than 20% of patients suffering from additional autoimmune diseases aside from the illnesses for which treatment is received (Vista et al., 2017), the development of an expert system can greatly improve the state of healthcare in the country. This is especially valuable for rural communities; less than a quarter receive medical checkups (Reyes-Gibby & Aday, 2005), and these figures may be exacerbated due to limited awareness and knowledge on rarer autoimmune diseases.

KNOWLEDGE BASE AND ALGORITHM

This section gives a technical description of the system's knowledge base and underlying algorithm.

Symptoms. The symptoms for the diseases were curated from four primary technical resources: *(i)* the peer-reviewed MSD Manuals by the pharmaceutical company Merck & Co., Inc., alongside the disease databases of *(ii)* the academic medical center Cleveland Clinic, *(iii)* the National Institutes of Health of the United States, and *(iv)* the National Health Service of the United Kingdom. The final list was then verified by Joan Tan, M.D., a physician from St. Luke's Medical Center, Philippines.

Backward Chaining with Uncertainty. Medical diagnosis is an extremely nuanced task, even for human experts. One of the chief reasons is that, although symptoms are indicative of a disease, a disease itself cannot be unequivocally interpreted as a mere conjunction of all the symptoms or of a clear-cut subset of them. In this regard, the problem transposes into that of incorporating uncertainty.

Since machines operate on the basis of numerical values, it is necessary to assign each symptom a weight. The following heuristic was used by the system: $1 - \frac{\text{number of diseases with the symptom}}{\text{total number of diseases}}$. Essentially an ad-hoc estimate, it gives greater weight to symptoms that are more unique to a certain disease. For instance, bald patches are distinct to alopecia areata in the system's universe of discourse; thus, the weight assigned to this indicator is high: $1 - \frac{1}{15} = 0.9333$. On the flip side, it assigns lower weights to less discriminatory symptoms like fatigue ($1 - \frac{9}{15} = 0.4$) and weight loss: ($1 - \frac{8}{15} = 0.4667$).

Combined with the native backward-chaining mechanism of Prolog, this allows the expert system to admit uncertainty to a certain extent. Patterned after the rules pioneered by the backward-chaining expert system MYCIN, written in Lisp by Stanford University's Edward Shortliffe (Shortliffe, 1977), the certainty level (confidence factor) is adjusted based on the patient's response. Formally, suppose the certainty level after the i^{th} query is c_i and the weight of the symptom s_i is w_i. In light of sentential calculus, let ψ_i be the proposition "If the patient reports experiencing s_i, then he/she has this disease."

Initially, $c_0 \coloneqq 0$. If the symptom s_i is reported, then the sentence ψ_i is created. The conjunction of this sentence with the previous ones creates a new sentence $\psi_1 \land \psi_2 \land \ldots \land \psi_i$ with certainty level c_i, the computation of which is adapted from MYCIN's parallel combination of supplied evidence. The implementation of this piecewise computation in the logic language Prolog is shown on the next page (note that only the first match following Prolog's top-down search strategy is returned):

```
confidenceFactor(CF, TrueWeight, NewCF) :- CF > 0, TrueWeight > 0,
   NewCF is CF + TrueWeight - CF * TrueWeight.
confidenceFactor(CF, TrueWeight, NewCF) :- CF < 0, TrueWeight < 0,
   NewCF is CF + TrueWeight + CF * TrueWeight.
confidenceFactor(CF, TrueWeight, NewCF) :- NewCF is (CF + TrueWeight)
   / (1 - min(abs(CF), abs(TrueWeight))).
```

On the other hand, if the patient gives a negative reply to experiencing the symptom s_i, its weight is taken to be its negative, i.e., $w_i \coloneqq -w_i$, representing a decrease in the certainty level of the agent. Nevertheless, the computation for the confidence factor is unaltered. The Prolog rule for this is as follows (`Answer` is set to 1 if the patient's response is affirmative or to 0 if the response is negative):

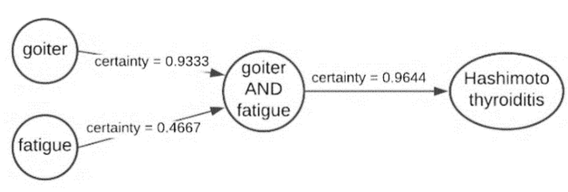
```
adjustedWeight(Weight, Answer, NewWeight) :- Answer =@= 1, NewWeight
   is Weight.
adjustedWeight(Weight, Answer, NewWeight) :- Answer =@= 0, NewWeight
   is -1 * Weight.
```

Diagnosis. To avoid redundancy and reduce the length of the probing, a disease is ruled out once the certainty level falls below 0.2 (following MYCIN's threshold). It then proceeds to the next disease. In Prolog, `dynamic/1` predicates for the list of symptoms per disease allowed for run-time exclusion of already-queried symptoms via a sequence of `delete/3`, `retract/1`, and `asserta/1`. In declarative fashion, a list is also maintained for the certainty levels corresponding to the diseases. A fact, `mapping(Index, Disease)`, was created to match each disease with the pertinent list element.

Typically, the disease with which the system records the highest certainty level is reported as the diagnosis. However, if a disease registers a certainty level of at least 0.9, a diagnosis is immediately given, thus terminating the inquiry process. If none of the diseases in the knowledge base satisfy the 0.2 threshold, then the patient is referred to a larger medical facility for a more thorough diagnosis. However, if a symptom indicative of an emergency is reported, this is overridden, and a diagnosis is compulsorily given with the confidence factor. In the knowledge base, there are three such cases: chest pain, abnormally slow heart rate (below 30 beats per minute), and body temperature above 40 °C.

User Interface. Since the target users of this medical expert system include the patient and an accompanying health worker, probing is classified into two veins: those that do not involve medical expertise (e.g., asking a patient if he/she has difficulty breathing) and those that require specialized knowledge (e.g., asking a health care worker if the patient shows Gottron papules on their hand joints). The prompts are stored in the knowledge base as a collection of `display(Symptom, Prompt)` facts. Finally, the bidirectional library JPL was imported to connect the knowledge base to the graphical user interface through the SWI Prolog Foreign Language Interface and the Java Native Interface.

Difficulties Faced. The first phase, knowledge gathering, was generally straightforward due to the abundance of medical references that enumerate main symptoms for the immunological diseases. However, translating the multifaceted nature of medical diagnosis into a set of logical rules was a focal challenge. Even physicians interviewed admitted the difficulty of reducing diagnosis to an algorithmic sequence, as human expertise in probing and evaluation does not readily structure itself to a procedure.



Although the knowledge base is greatly simplified into a specific niche, it is imperative to consider that, while symptoms are indicators of diseases, not all of these have to be present for a diagnosis to be given. Several approaches were assessed during development. For instance, a majority rule may require the patient to exhibit only 50% +1 of the symptoms. The weakness is that it hinges on the assumption of the list in the knowledge base being fully exhaustive and does not take into account that some symptoms are more discriminatory. The pair, thus, reviewed best practices employed by existing medical expert systems and found that mapping a symptom to some form of numerical weight allows a MYCIN-style calculation of the certainty factor to be performed as illustrated in the adjoining diagram.

EVALUATION AND ANALYSIS

The strength of the system is its ability to capture the calculus of uncertainty that characterizes actual medical diagnosis with its calculation of confidence levels, markedly requiring fewer simplifying assumptions. Since each symptom is assigned a weight, it is less susceptible to false negatives against less sophisticated approaches. For example, the gravity of denying joint pain is lower than that of denying goiter. Analogously, reporting goiter significantly raises the certainty of this thyroiditis. Using Prolog's native backward chaining allowed for a more robust implementation of these rules.

As a consultative system designed for patients and healthcare workers, having a graphical user interface, vis-à-vis Filipino translations of the questions, makes it more intuitive. Clinical investigations are refined and made more thorough with the inclusion of technical indicators, such as mucocutaneous lesions, and emergencies are alerted with a pop-up message. The system is also capable of tailoring the list of symptoms to the background of the patient (i.e., failure to thrive is removed for adults, and vaginal dryness and irregular menstrual periods are removed for male patients). Finally, the qualitative description for the confidence intervals facilitates a user-centered reporting of the final diagnosis.

Nonetheless, it is still subject to the ***inherent limitations*** of intelligent systems. Foremost is the closed-world assumption. Although the system sought to mitigate it with carefully worded semantics to dampen any definitive conclusion, the limited size of the knowledge base, along with the fact that only clinical manifestations (without laboratory tests) can be utilized to establish results, made it more prone to aggressive diagnosis and false positives. For instance, a patient reporting hair loss, multiple bald patches, and itchiness on these bald spots will be mapped with probable alopecia areata. However, it is possible that these symptoms point out to another alopecia-related condition outside the scope of autoimmune disorders; in actual practice, trichoscopy (involving microscopy) has to be conducted.

The second area is in the incorporation of uncertainty into the model. Most statements found in medical literature and given by physicians are generally restricted to quasi-quantitative descriptors such as "frequent" or "general" symptoms (instead of giving exact numerical or ordinal ranking). Due to this bottleneck on knowledge acquisition, the pair had to formulate an ad-hoc heuristic for the weight, subject to the closed-world assumption. The methodological robustness of MYCIN has also drawn scrutiny; it also employs ad-hoc formulae that may not be readily generalizable to other applications.

Heckerman (1992) also proved that they hinge on the same Bayesian assumption (independence of conditions), which is unlikely for medical symptoms. For instance, goiter *causes* fatigue, but this link is overlooked in the parallel combination of evidence for Hashimoto thyroiditis. Although a necessary trade-off to simplify complex disease dynamics, the need for models closer to the real world is justified.

SUMMARY AND LESSONS LEARNED

In summary, EHRLICH is an expert system developed to automate the diagnosis of 15 different autoimmune diseases. Central to its design and function are Prolog's backward-chaining mechanism, augmented by the incorporation of ad hoc metrics to use with MYCIN's certainty computations. Though the implementation of certainty contributed to the increased reliability of the system, it is still subject to limitations that, for now, prevent the widespread adoption of similar systems in needy areas.

In contrast to the systems often developed through an imperative paradigm, knowledge engineering requires intimate connections among its component fields; the development of this system involved incorporating concepts from medical diagnosis, statistics, and algorithm design, with the facets of one component building from the others. While accuracy and reliability can be improved upon within the scope of the expert system, the closed-world assumption makes it difficult to offer credible generalizations for issues outside its expertise. Finally, while doctors regard symptoms as results of a disease, the system conversely diagnoses diseases by treating them as clusters of symptoms, an interesting note on how knowledge engineering offers another approach to the usual epistemology.

Heckerman, D. (1992). The certainty-factor model. In *Encyclopedia of Artificial Intelligence,* 131-138.
Reyes-Gibby, C.C., & Abay, L.A. (2005). Access to health care in a rural area of the Philippines. Ethnicity & Disease, *15*(1), 104-109.
Shortliffe, E.H. (1977). MYCIN: A knowledge-based computer program applied to infectious diseases. *Proceedings of the Annual Symposium on Computer Application in Medical Healthcare*, 66-69.
Vista, E.S., Weisman, M.H., Ishimori, M.L., Chen, H., Bourn, R.L., Bruner, B.F., Hamiyojo, L., Tanangunan, R.D., Gal, N.J., Robertson, J.M., Harley, J.B., Guthridge, J.M., Navarra, S.V., & James, J.A. (2017). String viral associations with SLE among Filipinos. Lupus Science and Medicine, *4*(1).