A Dual Inference Expert System Using Forward and Backward Chaining for Early Screening of Bone Disease

Nikmasari Pakaya[1*], Lillyan Hadjaratie[2], Nazri Pri Andini Zulqa Dah Hunawa[3], Moh. Hidayat Koniyo[4], Indhitya R. Padiku[5], Rahmat Taufik R. L. Bau[6]

[1,2,3,4,5,6]Information System, Faculty of Engineering, Universitas Negeri Gorontalo, Gorontalo, Indonesia
E-mail: [1]nikmasari.pakaya@ung.ac.id, [2]lillyan.hadjaratie@ung.ac.id, [3]nazrihunawa03@gmail.com, [4]hidayat_koniyo@ung.ac.id, [5]indypadiku@ung.ac.id, [6]rahmattaufik@ung.ac.id

(Received: 26 Sep 2026, revised: 11 Oct 2026, accepted: 16 Oct 2026)

Abstract

Bone diseases can affect mobility and quality of life, while limited access to healthcare services may hinder the early identification of related symptoms. This study aimed to develop a dual-inference expert system for the early screening of bone diseases using forward chaining and backward chaining. The system was implemented as a web-based application that allows users to select either inference method within a single application and shared knowledge base. The knowledge base comprises six bone diseases, 33 unique symptoms, and six IF–THEN rules. Evaluation involved expert assessment using Aiken's V, black-box testing, and the System Usability Scale (SUS). The overall Aiken's V was 0.917 for 34 symptom–disease associations and 0.856 for IF–THEN rule structures. The mean Aiken's V for the use of the AND operator was 0.722, indicating that the rule conditions require further examination. Black-box testing verified the functional operation of the implemented features, while usability evaluation involving 25 respondents yielded an average SUS score of 74.3. These results provide evidence of expert-assessed content validity, functional operation, and favorable perceived usability, but they do not establish clinical diagnostic accuracy. The system may serve as a preliminary screening aid to help users recognize possible bone-related conditions and consider seeking professional medical evaluation.

Keywords: Expert System, Forward Chaining, Backward Chaining, Bone Disease, Early Screening, Usability Evaluation.

I. INTRODUCTION

Bone-related diseases represent a significant public health concern because they can impair mobility, physical functioning, and quality of life. The progressive nature of several bone and musculoskeletal conditions also makes early recognition important, particularly when initial symptoms are often perceived as common physical complaints. In Indonesia, osteoarthritis has become an increasing health burden, with the number of cases more than doubling between 1990 and 2019 and a substantial increase in years lived with disability (YLD) [1], [2]. These conditions emphasize the importance of accessible preliminary screening that can assist individuals in recognizing symptoms that may require further medical attention.

The burden of bone-related conditions is further reflected in the increasing incidence of fractures and the prevalence of chronic bone diseases worldwide. In 2019, approximately 178 million new fracture cases were reported globally, representing a 33.4% increase compared with 1990, while approximately 455 million prevalent cases associated with acute and chronic fracture symptoms were recorded, an increase of 70.1% over the same period [3]. Osteoporosis alone is estimated to affect approximately 200 million people worldwide [4]. In Indonesia, the 2018 Basic Health Research reported an osteoarthritis prevalence of 7.30% among individuals aged 15 years and older, while Gorontalo Province recorded a prevalence of 6.85% [5], [6]. The continuing burden of these conditions indicates the need for approaches that facilitate early recognition before individuals proceed to more comprehensive clinical assessment.

The need for accessible early screening is particularly relevant in Gorontalo, where the availability of orthopedic specialists may be limited relative to the population. Data from Indonesia's Ministry of Health list three orthopedic and traumatology specialists for a population of approximately 1.26 million in the province [7]. In addition, self-medication has been identified as one of the reasons why individuals with health complaints may not seek outpatient care, highlighting a potential barrier to obtaining professional assessment [8]. This concern is consistent with evidence of the uneven distribution of medical specialists across districts in Indonesia

DOI: 10.34148/teknika.v15i3.1609

[9]. In this context, an accessible information and preliminary screening mechanism may help users recognize symptoms that warrant further medical assessment. Such a mechanism should complement, rather than replace, professional examination and diagnosis.

Advances in information technology and artificial intelligence provide opportunities to address this need through expert systems. A knowledge-based expert system represents domain knowledge using structured rules and applies an inference mechanism to generate conclusions based on available facts or hypotheses [10], [11]. In healthcare, expert systems have been applied to support preliminary diagnosis and symptom interpretation [12][13]. These studies demonstrate the potential of rule-based expert systems to provide accessible decision-support functions, but the effectiveness of such systems also depends on how the inference mechanism corresponds to the reasoning situation faced by users.

Two commonly used rule-based inference mechanisms are Forward Chaining and Backward Chaining, which operate in different directions. Forward Chaining is fact-driven or symptom-driven: the reasoning process begins with facts supplied by the user and proceeds toward possible conclusions. Backward Chaining is goal-driven or hypothesis-driven: the reasoning process begins with a suspected conclusion and proceeds backward to verify whether the supporting facts or symptoms are present. Previous research has applied Backward Chaining in medical expert systems [13], Forward Chaining for early diagnosis of bone disease [14], and Forward Chaining combined with a certainty factor approach for symptom-based diagnosis [15]. Although these studies demonstrate the applicability of rule-based inference in medical domains, the existing approaches generally employ one principal inference direction or combine Forward Chaining with a certainty mechanism. The use of two alternative inference directions for preliminary screening of multiple bone diseases within a single application remains limited.

This limitation constitutes the main methodological and practical gap addressed in this study. In preliminary bone disease screening, users may approach the system from different reasoning situations. A user who has experienced several symptoms may want the system to identify possible diseases, which is naturally supported by Forward Chaining. Conversely, a user who wants to examine a particular suspected disease may prefer to start from that disease hypothesis and verify its associated symptoms, which is more naturally supported by Backward Chaining. A system restricted to only one inference direction cannot directly accommodate both situations within the same application. Therefore, rather than developing two separate applications or claiming a new inference algorithm, this study positions a dual-inference expert system as a single screening application that provides two selectable inference modes: Forward Chaining and Backward Chaining.

The proposed system consequently treats Forward Chaining and Backward Chaining as complementary but independent inference options within one application. The user selects the inference mode according to the screening approach they wish to use. In the Forward Chaining mode, the user provides symptoms and the system traces the corresponding rules to identify candidate bone diseases. In the Backward Chaining mode, the user selects or examines a disease hypothesis and the system traces backward to verify the symptoms associated with that hypothesis. Both modes use the same domain knowledge base covering six bone-related conditions, namely osteoporosis, osteoarthritis, osteomyelitis, rheumatic fever, vertebral fracture, and tuberculous spondylitis. The knowledge base was formulated from relevant medical literature and refined through consultation with orthopedic specialists. Thus, the dual-inference concept in this study refers specifically to the availability of two selectable reasoning mechanisms within one web-based application, not to a sequential or tightly integrated hybrid inference process.

Based on the identified problem and research gap, the objective of this study is to develop and evaluate a web-based dual-inference expert system for early screening of bone disease that provides users with a choice between Forward Chaining and Backward Chaining inference. The development is intended to address the need for an accessible preliminary screening tool in a context where access to orthopedic services is constrained, while allowing users to approach screening according to either symptom-driven or hypothesis-driven reasoning. The system is developed using a prototyping approach to support iterative refinement of requirements, knowledge representation, inference processes, and user interaction. Functional performance is evaluated using black-box testing, while user acceptability is assessed using the System Usability Scale (SUS).

The novelty of this study lies in the design and implementation of a single web-based expert system that offers two alternative and selectable rule-based inference modes for early screening of multiple bone diseases. Rather than proposing a new inference algorithm or sequentially combining Forward and Backward Chaining into one reasoning chain, the study demonstrates how the two inference directions can be made available within the same application to accommodate different user reasoning needs. This positioning provides a clear distinction from previous studies that focus primarily on a single inference direction or on Forward Chaining combined with certainty-based reasoning. The resulting system is intended as an accessible preliminary screening aid that can help users recognize possible bone-related conditions and determine the need for further medical consultation, while final diagnosis and treatment remain the responsibility of qualified healthcare professionals.

II. RESEARCH METHODOLOGY

This study employed a prototyping-based development approach to design and implement a web-based dual-inference expert system for the early screening of bone disease. The system was developed as a web application that provides two selectable inference modes: Forward Chaining

and Backward Chaining. Both inference modes use the same knowledge base but apply different reasoning directions. Forward Chaining begins with symptoms confirmed by the user and derives candidate diseases, whereas Backward Chaining begins with a disease hypothesis selected by the user and verifies the symptoms supporting that hypothesis. Thus, the two inference methods are implemented as alternative inference modes within one application rather than as sequential inference processes.

The research procedure consisted of a preliminary study followed by six prototype development stages (Figure 1): Communication, Quick Plan, Modeling/Quick Design, Construction of Prototype, Deployment, and Delivery & Feedback [16]. The preliminary study established the research problem, screening scope, disease domain, and initial information requirements. These outputs became the basis for the Communication stage and subsequent system development.

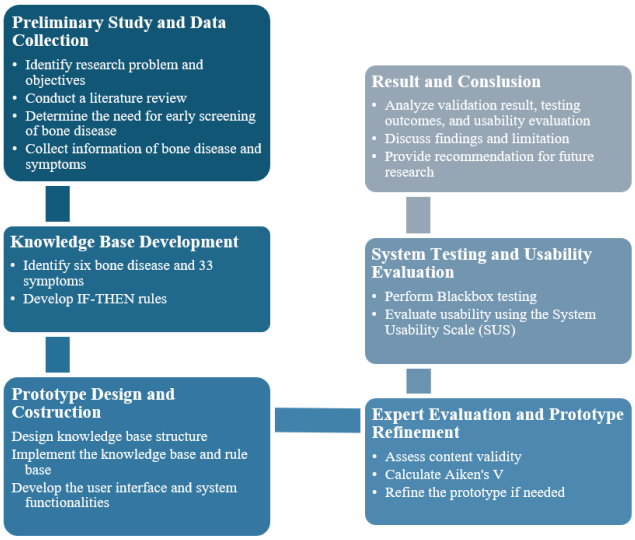


Figure 1. Research Stages

A. Preliminary Study and Data Collection

The preliminary study was conducted to establish the problem domain, the scope of the early bone disease screening system, and identify the information required for knowledge base construction and system development. Data were obtained through three complementary activities: interviews, observation, and literature study.

Interviews were conducted with three orthopedic specialists to obtain domain knowledge concerning the selected bone diseases, associated symptoms, and relationships between symptoms and diseases. All three specialists had equivalent roles in the knowledge acquisition, review, refinement, and validation processes. The information obtained from the three experts was used together with findings from the literature study to identify the diseases and symptoms included in the screening scope. The experts also reviewed the disease–symptom relationships and the rules formulated from the collected information. Their input was used to refine the knowledge representation before it was implemented in the expert system.

Observation was used to obtain information related to the intended use of the system and the requirements of the early screening process. The observation focused on the user interaction requirements and the practical need for obtaining an early screening of bone related complaints. The observation results were used to complement information obtained from expert consultation and literature study, particularly in defining the functional requirements and interaction flow of the web-based application.

The literature study was conducted to identify bone diseases and their symptoms, and previously reported applications of expert systems and inference methods. Medical and scientific references were examined to establish the initial disease and symptom concepts and to support the identification of disease–symptom relationships. The information obtained from the literature was not directly used as final rules. Instead, it was organized into an initial knowledge and subsequently reviewed by the three orthopedic specialists.

B. Knowledge Base Construction

The knowledge base was constructed through four main activities: disease identification, symptom identification, disease–symptom relationship construction, and knowledge representation. The final knowledge base consists of six bone diseases and 33 coded symptoms, represented by G01–G33. The six diseases are Osteoporosis (P01), Osteoarthritis (P02), Osteomyelitis (P03), Rheumatic Fever (P04), Vertebral Fracture (P05), and Tuberculous Spondylitis (P06). The six disease codes are used consistently throughout the knowledge base and inference processes.

Symptom knowledge consists of the symptoms associated with the six selected diseases. Each symptom is assigned a unique code from G01 to G33 to facilitate data processing and rule representation. The final knowledge base contains 33 unique coded symptoms and was refined through expert review before being implemented in the early screening system.

The disease–symptom relationships were established by mapping each coded symptom to the disease or diseases with which it was associated. The resulting relationships were reviewed by the three orthopedic specialists and subsequently translated into inference rules. The rule base consists of six disease-specific rules, which are outlined in Table 1.

Table 1. Rule Base

Rule P01 (Osteoporosis)	**IF** G01 **Pain in the back or spinal bones** **AND** G02 **Stooped posture** **AND** G03 **Decreased body height** **AND** G04 **Bones fracture easily, even after minor impacts** **AND** G05 **History of menopause** **AND** G06 **Patient is aged approximately 48–51 years or older** **THEN** ***OSTEOPOROSIS***
Rule P02 (Osteoarthritis)	IF **G07** Joint pain and swelling

DOI: 10.34148/teknika.v15i3.1609
TEKNIKA, Volume 15(3), November 2026, pp. 541-552
ISSN 2549-8037, EISSN 2549-8045

	AND **G08** Joint stiffness in the morning lasting 30 minutes or occurring after a prolonged period of rest AND **G09** Joint sounds during movement (crepitus, or a "crackling" sound) AND **G10** Pain that increases when the joint is moved THEN *OSTEOARTHRITIS*
Rule P03 (Osteomyelitis)	IF **G11** History of injury or surgery AND **G12** Swelling and redness in the infected area AND **G13** Fever AND **G14** General weakness, fatigue, or feeling unwell AND **G15** Pain when pressure is applied to the affected area AND **G16** A persistent, non-healing boil in the bone area THEN *OSTEOMYELITIS*
Rule P04 (Rheumatic Fever)	IF **G13** Fever AND **G17** Joint swelling, redness, and pain, particularly in the elbows, knees, wrists, and ankles AND **G18** Joint pain that may spread to other joints AND **G19** Red skin rash THEN *RHEUMATIC FEVER*
Rule P05 (Vertebral Fracture)	IF **G20** Pain along the spine, from the neck to the tailbone AND **G21** Tingling or numbness AND **G22** Curved or hunched spinal posture AND **G23** Involuntary muscle spasms AND **G24** Decreased body height AND **G25** Pinched nerve symptoms, such as weakness or paralysis THEN *VERTEBRAL FRACTURE*
Rule P06 (Tuberculous Spondylitis)	IF **G26** History of contact with a person with tuberculosis (TB) AND **G27** Back pain AND **G28** Neurological symptoms, such as leg weakness AND **G29** Excessive fatigue AND **G30** Spinal stiffness AND **G31** A lump or swelling in the back area AND **G32** Loss of appetite and weight loss AND **G33** Frequent night sweats THEN *TUBERCULOUS SPONDYLITIS (SPONDYLITIS TB)*

The relationships show that one symptom may be associated with more than one disease. For example, G13 (fever) occurs in both the Osteomyelitis and Rheumatic Fever rules. Such overlapping symptoms are retained because they represent relationships established in the knowledge base and are subsequently processed by the inference mechanism.

C. Knowledge Base Validation and Consistency Checking

Knowledge base validation was conducted after the disease–symptom relationships and rules had been formulated. The three orthopedic specialists reviewed the disease list, symptom list, disease–symptom relationships, and resulting IF–THEN rules. All three experts had equivalent roles in this process. The validation process focused on determining whether the represented symptoms were relevant to the selected diseases and whether the relationships encoded in the rules were consistent with the intended preliminary screening domain. Inconsistencies identified during the review were revised before implementation.

Consistency checking was also conducted at the representation and implementation levels. Each symptom used in an inference rule was required to have a unique G-code and to be available as an input to the inference process. Similarly, each disease included in the screening scope was required to have a corresponding disease code and rule. This checking resulted in a final knowledge base containing 33 coded symptoms and six inference rules. The previously listed "Limited range of motion" entry for Osteoarthritis was removed because it did not satisfy the coding and rule-representation requirements.

The content validity of the final rule representation was subsequently assessed through expert validation using Aiken's V [17] . This validation was intended to assess the validity of the knowledge representation based on expert judgment and was not interpreted as clinical diagnostic accuracy validation.

D. Inference Mechanism

Forward Chaining

Forward Chaining starts from symptoms confirmed by the user and proceeds toward candidate diseases. Initially, the system considers the six diseases in the knowledge base as possible candidates. The system then evaluates the confirmed symptoms against the disease–symptom relationships. To determine informative symptoms during the inference process, the frequency of a symptom among the remaining candidate diseases can be represented as Equation 1:

$$f(G_i) = \text{number of diseases that have symptom } G_i \quad (1)$$

where:

$f(G_i)$: Frequency or number of diseases that have the symptom i

G_i : The symptom i that is still relevant to the candidate disease

For each confirmed symptom, the candidate disease set is updated according to its relationship with the symptom. If a confirmed symptom is associated with a disease, that disease remains a candidate. When the inference iteration is completed, the system calculates the matching score for each remaining disease using Equation 2:

$$Score_{FC}(P_i) = \frac{C(P_i)}{T(P_i)} \, x \, 100\% \quad (2)$$

where:

$Score_{FC}(P_i)$: the symptom-to-disease match score for disease P_i using the Forward Chaining method.

$C(P_i)$: the number of symptoms confirmed by the user that match the symptoms in the rule for disease P_i

$T(P_i)$: the total number of symptoms listed in the rules for disease P_i

The disease with the highest matching score is selected as the preliminary screening result using Equation 3:

$$P_{best} = \underset{P_i \in P}{arg\,max}\ Score_{FC}\ P_i \quad (3)$$

where:
P_{best} : the disease with the highest match score
P : the set of all diseases evaluated by the system
P_i : the i-th disease
$Score_{FC}$: the symptom matching score for disease i
$arg\ max$: the operator that selects the disease identity with the maximum score value, rather than merely the highest score value.

Backward Chaining

Backward Chaining starts from a disease hypothesis selected by the user and works backward to verify the symptoms required by that hypothesis. In contrast to Forward Chaining, which starts from available symptoms, Backward Chaining starts from the selected goal. After the user selects a disease hypothesis, the system retrieves the corresponding rule and presents the required symptoms as sub-goals to be verified. Each symptom is encoded using a binary value using Equation 4:

$$x_{ij} = \begin{cases} 1,\ if\ G_j\ is\ confirmed \\ 0, if\ G_j\ is\ not\ confirmed \end{cases} \quad (4)$$

where:
X_{ij}: binary value indicating the confirmation status of symptom $G_j$for disease candidate P_i
G_j: the j-th symptom
P_i : the i-th disease candidate being evaluated
i : index of the disease candidate
j : index of the symptom
1 : the symptom is confirmed by the user
0 : the symptom is not confirmed by the user

The binary responses are then used to calculate the support score using Equation 5:

$$Score_{BC}(P_i) = \frac{\sum_{j=1}^{k_i} x_{ij}}{k_i}\ x\ 100\% \quad (5)$$

where:
$Score_{BC}(P_i)$: the symptom-matching score for disease candidate P_i using the Backward Chaining method
P_i : the i-th disease candidate selected as a hypothesis for evaluation
x_{ij} : a binary value indicating whether symptom $G_j$is confirmed for disease candidate P_i
k_i : the total number of symptoms associated with disease candidate $P_i$in the corresponding rule
j : the index of a symptom associated with disease candidate P_i

The resulting value indicates the degree of symptom support for the selected hypothesis.

Although both inference modes utilize the same knowledge base—comprising 33 symptoms and six rules—they differ in their starting points and directions of reasoning, as outlined in Table 2.

Table 2. Comparison of Forward and Backward Chaining

Aspect	Forward Chaining	Backward Chaining
Starting point	User-confirmed symptoms	Disease hypothesis selected by the user
Reasoning direction	Symptoms → disease candidates	Disease hypothesis → supporting symptoms
Main process	Matches confirmed symptoms with candidate diseases	Verifies symptoms required by the selected hypothesis
Knowledge base	Same	Same
Rule representation	IF symptoms THEN disease	Same rule evaluated backward from the disease goal
Quantitative output	Matching score	Support score
Output interpretation	Early screening result	Early screening result

The comparison does not represent two independent expert systems. Instead, it evaluates two reasoning approaches implemented as selectable modes in the same dual-inference system.

E. Prototype Development

The system was developed using a six-stage prototyping approach: Communication, Quick Plan, Modeling/Quick Design, Construction of Prototype, Deployment, and Delivery & Feedback. Each stage produced outputs that informed the subsequent stage, while feedback could trigger refinement of an earlier stage when requirements, knowledge representation, or functionality were incomplete or inconsistent.

Communication

The Communication stage focused on identifying system requirements and acquiring domain knowledge. The outputs of the preliminary study, literature study, expert consultation, observation, and identification of user requirements were consolidated at this stage. The outputs were the defined early screening scope, disease and symptom requirements, knowledge base, inference mode requirements, user interaction requirements, and system functional requirements.

Quick Plan

The Quick Plan stage translated the identified requirements into the system structure and functional modules. The main modules included symptom input, selection of the inference mode, knowledge base

management, inference processing, early screening result generation, and screening history storage. The planning stage also established that Forward Chaining and Backward Chaining would be implemented in a dual inference Expert System.

Modeling/Quick Design

The Modeling/Quick Design stage translated the requirements into system models and interface designs. The design included the context diagram, external-entity model, data flow diagrams (DFD), database structure and user interface prototypes. The Forward Chaining interface was designed around symptom-driven inference, whereas the Backward Chaining interface was designed around the selection of a disease hypothesis.

Construction of Prototype

During the Construction of Prototype stage, the system design was translated into a functional web-based system. The system was developed using PHP with the Laravel framework, while MySQL was used to manage user inputs, disease and symptom data, inference rules, and screening histories. The knowledge base, consisting of six diseases, 33 coded symptoms, and six rules, was integrated into the system. Both Forward Chaining and Backward Chaining were implemented using the same knowledge base as inference modes within a single system.

Deployment

The deployment phase involves making the prototype available for evaluation within the intended initial screening process. The deployed system provides the necessary functions for symptom input, inference mode selection, inference execution, presentation of initial screening results, and storage of screening history. The system resulting from this phase is subsequently used for evaluation and user feedback collection during the Delivery & Feedback phase.

Delivery & Feedback

The Delivery & Feedback stage gathers responses regarding the use of the prototype and identifies aspects requiring refinement. Feedback concerning usability and system functionality is used to determine whether modifications are necessary. If the feedback reveals incompleteness or inconsistencies in system requirements or functionality, the prototype may be revisited. This iterative mechanism ensures that the prototype can be refined based on evaluation results and feedback.

F. System Testing and Evaluation

System evaluation was conducted from three complementary perspectives: functional testing using black-box testing, expert validation using Aiken's V, and usability evaluation using the System Usability Scale (SUS). Black-box testing was conducted to verify the functional system behavior without examining its internal source-code implementation. The tested functions included symptom input, inference-mode selection, execution of Forward Chaining, execution of Backward Chaining, generation of early screening results, and storage of screening history. The test results were used to determine whether the implemented functions operated according to the specified requirements.

Black-box testing was used as functional verification only. It was not used to calculate diagnostic accuracy and should not be interpreted as evidence of clinical validity. Expert validation was conducted by the three orthopedic specialists who were involved throughout the knowledge acquisition, review, refinement, and validation processes. The resulting expert ratings were analyzed using Aiken's V to quantify the content validity of the assessed items. Aiken's V can be expressed as Equation 6:

$$V = \frac{\sum_{i=1}^{n} s_i}{n(c-1)} \tag{6}$$

where:

V : Aiken's V content validity coefficient
$s_i = r_i - l_o$: the score assigned by the i-th expert after subtracting the lowest possible rating
r_i : the rating assigned by the i-th expert
l_o : the lowest possible rating on the rating scale
n : the number of experts or raters
c : the highest possible rating on the rating scale

The Aiken's V analysis provides quantitative evidence of the agreement of expert judgments regarding the content validity of the assessed knowledge representation. It does not constitute clinical diagnostic accuracy testing against physician-confirmed patient diagnoses.

Usability was evaluated using the System Usability Scale (SUS), a standardized questionnaire comprising ten statements regarding perceived utility and ease of use. The usability evaluation involved 25 respondents recruited via convenience sampling, based on ease of access and their willingness to participate. The group comprised two physicians, two administrative staff members, and 21 members of the general public (patients) across three different medical clinics. The system was first tested by the physicians, followed by the other respondents. All 25 respondents interacted with the web-based early screening system for bone disease before completing the System Usability Scale (SUS) questionnaire. This questionnaire was used to assess respondents' perceptions of the system's usability based on their direct experience using the application. The resulting SUS scores reflect perceptions of system usability and do not determine the system's clinical diagnostic accuracy.

III. RESULTS AND DISCUSSION

Result

Expert validation using Aiken's V

The content validity assessment involved three experts who evaluated the relevance of symptom disease relationships using a five-point rating scale. Aiken's V was calculated for each item using a minimum score of 1, a maximum score of 5. The knowledge base consists of 33 unique symptoms. During content validation, 34 symptom–disease associations

were assessed because one symptom was associated with more than one disease. Aiken's V was calculated for each association to evaluate its relevance to the corresponding disease. The assessment covered 34 symptom–disease relationships involving six bone-related conditions. As shown in Table 3, the overall mean Aiken's V coefficient for symptom–disease relationships was 0.917. As presented in Table 3, the highest mean was obtained for osteoarthritis (V = 0.958), followed by tuberculous spondylitis (V = 0.948), osteoporosis (V = 0.944), vertebral fracture (V = 0.931), rheumatic fever (V = 0.896), and osteomyelitis (V = 0.819).

Table 3. Aiken's V for Symptom–Disease Relationships

Disease	Items (n)	Mean Aiken's V	Minimum V
Osteoporosis	6	0.944	0.750
Osteoarthritis	4	0.958	0.833
Osteomyelitis	6	0.819	0.750
Rheumatic fever	4	0.896	0.750
Vertebral fracture	6	0.931	0.750
Tuberculous Spondylitis	8	0.948	0.750
Overall	**34**	**0.917**	**0.750**

The item analysis identified 20 of the 34 relationships with an Aiken's V of 1.000, indicating that all three experts assigned the maximum rating of 5. The remaining 14 relationships obtained coefficients between 0.750 and 0.833. Seven items received V = 0.750: G06 for osteoporosis; G12, G13, and G14 for osteomyelitis; G13 for rheumatic fever; G21 for vertebral fracture; and G29 for tuberculous spondylitis. These results provide an overview of the experts' ratings of symptom relevance.

Aiken's V was calculated for 18 items covering three dimensions: suitability of symptom combinations, appropriateness of the AND operator, and symptom adequacy and clarity. The overall mean Aiken's V for rule structures was 0.856. Table 4 presents the mean coefficient for each disease, calculated from its three assessment items.

Table 4. Aiken's V for IF–THEN Rule Structures

Disease	Symptom combination	AND operator	Symptom adequacy and clarity	Mean V
Osteoporosis	1.000	0.750	0.917	0.889
Osteoarthritis	1.000	0.750	0.750	0.833
Osteomyelitis	1.000	0.750	0.667	0.806
Rheumatic fever	1.000	0.667	1.000	0.889
Vertebral fracture	1.000	0.750	0.750	0.833
Tuberculous Spondylitis	1.000	0.667	1.000	0.889
Overall Mean	**1.000**	**0.722**	**0.847**	**0.856**

The overall mean Aiken's V for IF–THEN rule structures was 0.856, with values ranging from 0.667 to 1.000. Suitability of symptom combinations received a mean of 1.000, indicating agreement among the experts on this dimension. In comparison, the AND operator achieved a mean of 0.722, while symptom adequacy and clarity achieved 0.847. The lowest individual value, 0.667, occurred for symptom adequacy and clarity in the osteomyelitis rule (P03), AND operator in the rheumatic fever rule (P04), and tuberculous spondylitis (P06).

Web-Based System Development

The developed prototype was implemented as a web-based expert system for the early screening of bone diseases. It provides two selectable inference modes, Forward Chaining and Backward Chaining, within a single system that shares the same knowledge base. Forward Chaining begins with user-reported symptoms and evaluates the applicable rules to identify potential diseases. In contrast, Backward Chaining begins with a selected disease hypothesis and verifies the symptoms associated with that hypothesis. This dual inference design allows users to choose an inference approach according to their screening needs. The system also provides a rule-management interface through which authorized medical experts can review and update symptom–disease relationships. These features support the implementation and management of the knowledge base, while the clinical accuracy of the screening results requires separate evaluation.

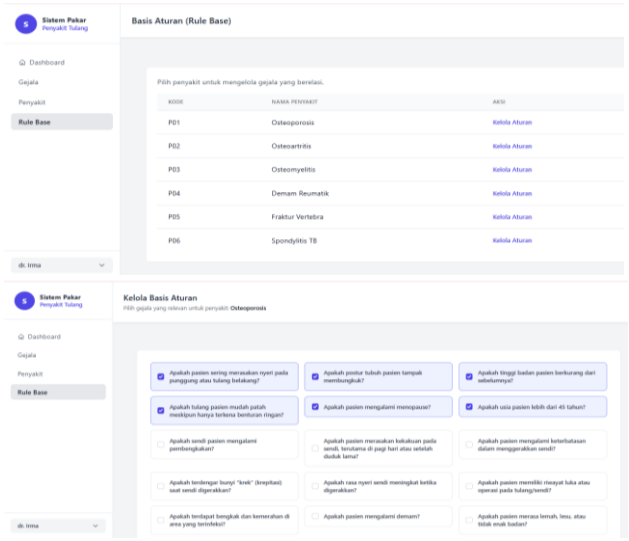


Figure 2. Selecting Rules with Checkboxes that Link Each Symptom to Each Disease

Figure 2 shows the symptom management rule feature, which can only be used by doctors. On this page, doctors can select rules using checkboxes that link each symptom to a specific disease. This makes it easier for doctors to define and update screening rules conveniently.

Meanwhile, on the patient's early screening page, shown in Figure 3, the system interactively asks if the patient is experiencing any symptoms. The patient can answer "Yes" if they are experiencing symptoms, or "No" if they are not experiencing symptoms. If the patient answers "Yes," the system will continue with the next relevant symptom question. This question-and-answer process continues until

DOI: 10.34148/teknika.v15i3.1609
TEKNIKA, Volume 15(3), November 2026, pp. 541-552
ISSN 2549-8037, EISSN 2549-8045

the system has collected enough data to display an appropriate early screening.

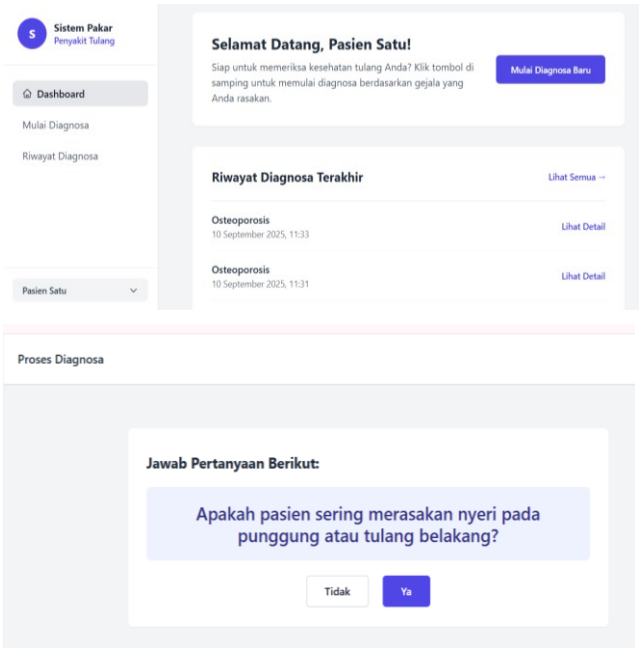


Figure 3. Forward Chaining Mode for Early Bone Disease Screening

In the Forward Chaining process, the system begins with symptoms confirmed by the user and compares them against disease rules stored in the knowledge base. In this illustrative case, the user confirms four out of the six symptoms associated with Osteoporosis (P01)—specifically G01, G02, G03, and G05—while G04 and G06 remain unconfirmed. The match score is calculated as:

$$Score_{FC}(P_1) = \frac{4}{6}\ x\ 100\% = 66.67\ \%$$

In Forward Chaining, the system performs similar comparisons against rules for other diseases to identify and rank potential disease candidates based on their match scores. Based on an analysis of symptom frequencies across six disease rules, the entered symptoms showed the highest match with rule P01 (Osteoporosis). The results indicate a match between the symptoms confirmed by the user and the rule associated with osteoporosis.

In many cases of bone disease, doctors, therapists, and even patients themselves often have initial suspicions based on the patient's age, risk factors, or existing complaints [18]. The backward chaining shown in Figure 4 enables the system to immediately test the suspected disease by verifying only the relevant symptoms. This replicates real-world clinical reasoning.

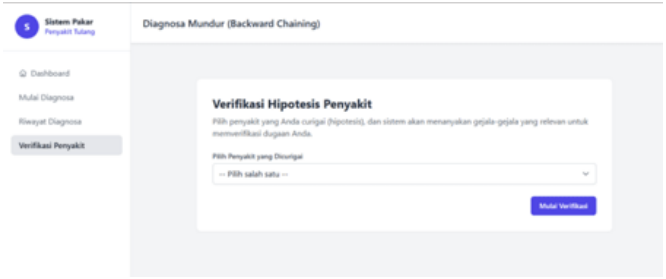


Figure 4. Backward Chaining Mode for Early Bone Disease Screening

In the Backward Chaining process, if the user selects Vertebral Fracture (P05) as the disease hypothesis to be evaluated, the system verifies six associated symptoms: G20, G21, G22, G23, G24, and G25. Suppose that, based on the responses provided, four symptoms are confirmed while the other two are not; the resulting support score is calculated as:

$$Score_{BC}(P_1) = \frac{4}{6}\ x\ 100\% = 66.67\ \%$$

This score represents the proportion of symptoms supporting the selected hypothesis—Vertebral Fracture (P05).

Black-box Testing Result

Black-box testing was conducted to evaluate whether the system functions according to its specified requirements. The test scenarios covered the execution of Forward Chaining and Backward Chaining, incomplete symptom information, cases in which no screening result could be generated, and storage of screening history. For each scenario, the actual system response was compared with the expected output and classified as passed or failed. The results were used to identify functional discrepancies and determine whether further corrections were required. These tests evaluated functional correctness and did not measure clinical diagnostic accuracy. Screening feature testing results are shown in Table 5.

Table 5. Black-box Testing Result

No	Test scenario	Expected output	Observation Result
1	Select symptoms and execute Forward Chaining.	The system processes the selected symptoms and displays the corresponding early screening result.	The symptoms selected are in accordance with the rules so that the early screening results are displayed. ***(Passed)***
2	Select a disease hypothesis and execute Backward Chaining.	The system verifies the symptoms associated with the selected hypothesis and displays the result according to the implemented inference logic.	The early screening results are displayed on the screen, showing disease findings that align with the hypothesis and the symptoms entered. ***(Passed)***

3	Provide incomplete symptom information.	The system follows its implemented logic for incomplete information, such as continuing relevant questions or indicating that the available information is insufficient	The system does not proceed with the inference process and instead displays a message indicating that the input information is incomplete. It offers the user the option to return to the previous screen to complete the information or to cancel the screening process. ***(Passed)***
4	Provide symptoms that do not satisfy any diagnostic rule	The system displays a message indicating that no early screening result can be generated.	The system displays a message stating that no disease matching the entered symptoms was found and advises the user to undergo further examination by an orthopedic specialist ***(Passed)***
5	Save a generated screening result.	The system stores the screening result in the screening history.	Diagnostic data is stored in the screening history. ***(Passed)***

System Usability Scale (SUS)

The system is presented to users to obtain feedback on its performance and suitability. This system usability assessment aims to identify parts of the system that do not meet user needs and expectations. evaluation activities were conducted using the System Usability Scale (SUS) [19] using with the following questions:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Table 6. System Usability Scale (SUS) Result

Respondent	Statement Score										Total Score	SUS Score
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10		
R1	5	1	5	1	5	1	5	1	5	1	40	100
R2	5	1	5	1	5	1	5	1	5	1	40	100
R3	4	2	4	4	4	3	5	2	4	4	26	65
R4	4	2	4	2	4	2	4	2	4	3	29	72.5
R5	4	3	4	2	5	2	4	2	4	3	29	72.5
R6	4	4	4	5	5	3	4	2	4	3	24	60
R7	4	2	4	4	4	2	5	2	4	4	27	67.5
R8	4	2	4	3	5	3	3	3	4	3	26	65
R9	4	2	5	3	5	2	5	2	5	3	32	80
R10	4	2	4	3	4	3	3	3	4	5	26	65
R11	4	2	4	2	4	2	4	2	4	4	28	70
R12	4	2	5	2	5	2	4	2	4	2	32	80
R13	4	3	4	2	5	2	3	2	4	3	28	70
R14	4	2	4	2	5	3	4	1	4	3	30	75
R15	4	2	5	2	4	2	4	2	4	3	30	75
R16	4	2	4	2	5	2	3	2	4	3	29	72.5
R17	4	2	4	2	4	2	4	2	5	3	30	75
R18	4	2	4	2	4	2	4	1	4	4	29	72.5
R19	4	2	5	3	5	2	5	2	4	3	31	77.5
R20	4	2	4	2	4	2	4	2	4	3	29	72.5
R21	4	2	4	4	4	2	4	2	4	4	28	70
R22	4	2	5	5	5	2	5	2	5	3	33	82.5
R23	4	2	4	4	5	2	5	1	5	2	34	85
R24	4	3	4	4	4	2	4	2	4	3	26	65
R25	4	2	4	4	4	2	4	2	3	4	27	67.5
Average Score												74.3

A questionnaire is distributed to 25 respondents to test the system's suitability and level of user satisfaction using the System Usability Scale (SUS) method. Respondents rate each statement using a Likert scale of 1 to 5. Table 6 in the file contains 25 SUS scores totaling 1,857.5. Thus:

$$\overline{\overline{SUS}} = \frac{\sum_{i=1}^{n} SUS_i}{n} = \frac{1857{,}5}{25} = 74.3$$

Based on the results, the system obtained a score of 74.3.

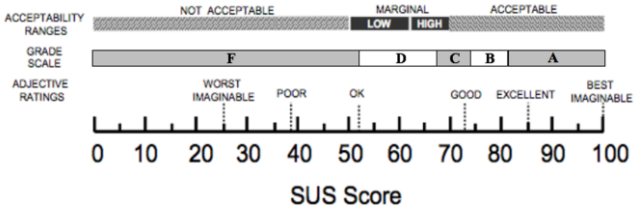


Figure 5. SUS Rating Scale [20]

This value exceeds the acceptability threshold of 68 based on the SUS rating scale shown in Figure 5; therefore, it can be concluded that the system possesses a good level of usability, is acceptable to users, and is viewed positively in terms of its usability. The result interpreted separately from the evaluation of the system's screening performance.

A. Discussion

This study developed a web-based dual inference expert system for early screening of bone disease using a prototyping approach. The main contribution is the implementation of two selectable inference modes, Forward Chaining and Backward Chaining, within a single application. Both modes use the same knowledge base but support different reasoning approaches. Forward Chaining starts from user-reported symptoms to identify potential bone diseases, whereas Backward Chaining starts from a selected disease hypothesis and verifies its associated symptoms. This design enables

users to select an inference approach according to their screening needs without requiring separate system.

The content validity assessment using Aiken's V examined two key components of the knowledge base: symptom–disease relationships and IF–THEN rule structures. The symptom–disease relationships obtained an overall mean Aiken's V value of 0.917, with values ranging from 0.750 to 1.000. At the disease level, osteoarthritis had the highest mean value (0.958), whereas osteomyelitis had the lowest (0.819). These differences indicate variation in the experts' assessments of symptom relevance across disease categories. Associations with lower values warrant further review to examine the specificity of the symptoms, the clarity of their descriptions, and their appropriateness for the corresponding diseases.

The IF–THEN structures obtained an overall mean Aiken's V value of 0.856. Among the three assessed dimensions, the suitability of symptom combinations had a mean value of 1.000, while the use of the AND operator had a mean value of 0.722. The adequacy and clarity of symptom descriptions obtained a mean value of 0.847. These results identify the logical structure of the rules as an important area for further examination. Requiring all symptoms in a rule to be present may restrict rule activation when symptoms do not always occur simultaneously. Therefore, the use of the AND operator should be reviewed in relation to the clinical characteristics of each disease. Symptom descriptions and symptom–disease associations with lower values should also be examined with medical experts. Any resulting revisions need to be applied consistently because both inference modes rely on the same knowledge base. Aiken's V provides evidence about expert-assessed content validity; it does not independently establish diagnostic accuracy or clinical effectiveness.

The black-box testing results provide evidence of the functional behavior of the implemented application. The testing scenarios cover symptom input, execution of both inference modes, incomplete symptom information, cases in which no screening result can be generated, and storage of screening history. These scenarios assess whether the application responds according to its functional requirements under the tested conditions. Nevertheless, successful functional testing demonstrates system behavior rather than the clinical correctness of its predictions.

The usability evaluation involving 25 respondents produced a mean System Usability Scale (SUS) score of 74.3. This score is above the benchmark of 68 and indicates a favorable perception of the system's usability. The finding concerns users' perceived interaction with the system and should be distinguished from the validity of the knowledge base and the functional performance assessed through black-box testing.

A significant limitation of this study is the absence of a diagnostic accuracy evaluation compared against an independent, physician-confirmed reference diagnosis. Nevertheless, the knowledge base was developed using medical literature and consultations with orthopedic specialists. Expert involvement in knowledge formulation and content assessment is not equivalent to clinical validation of the system's predictions. Future research should evaluate both inference modes using representative patient cases with established reference diagnoses. The system should be regarded as an early screening tool to assist users in identifying potential bone-related conditions, rather than as a substitute for professional diagnosis or treatment.

IV. CONCLUSION

This study developed a web-based dual-inference expert system for the early screening of bone diseases using a prototyping approach. The main contribution is the implementation of two selectable inference modes within a single application using a shared knowledge base comprising six bone diseases and 33 coded symptoms. Forward Chaining starts from user-reported symptoms and evaluates their relationships with potential diseases, whereas Backward Chaining starts from a user-selected disease hypothesis and verifies its associated symptoms. These alternative reasoning approaches allow users to interact with the system according to different screening needs without requiring separate applications.

During the expert evaluation, a medical specialist recommended clarifying the description of the "boil" symptom associated with osteomyelitis. The symptom statement (G16) was deemed insufficiently specific as it failed to describe the characteristics of the lesion. Based on this expert input, the description was refined to specify a persistent, non-healing boil in the bone area. This revision aimed to enhance the clarity of the symptom within the knowledge base. The revised description was confirmed by medical experts before being finalized and implemented in the prototype.

The expert assessment using Aiken's V yielded an overall mean coefficient of 0.917 for the 34 symptom–disease associations and 0.856 for the IF–THEN rule structures. The symptom–disease associations received generally favorable expert ratings, although some associations require further review. For the rule structures, the lower mean score for the AND operator (0.722), compared with symptom combination suitability (1.000) and symptom adequacy and clarity (0.847), highlights the need to reconsider whether all symptoms in each rule should be required simultaneously. These findings provide evidence of expert-assessed content validity for the knowledge base, but they do not establish the clinical diagnostic accuracy or effectiveness of the system. Black-box testing showed that the tested functions operated according to their specified requirements. Meanwhile, usability evaluation involving 25 respondents yielded a System Usability Scale (SUS) score of 74.3 indicating good perceived usability and user acceptability. These findings concern content assessment, functional performance, and usability; they do not establish the clinical diagnostic accuracy or effectiveness of the system.

A key limitation of this study is the absence of clinical validation against independent, physician-confirmed diagnoses. Future research should refine the knowledge base

based on further expert review, particularly the symptom–disease relationships and rule conditions requiring additional examination. The system should also be evaluated using representative clinical cases with appropriate reference diagnoses to measure diagnostic performance, including sensitivity, specificity, and accuracy for each disease where applicable. The system should be used only as an early screening aid to help users recognize possible bone-related conditions and consider seeking professional medical assessment, not as a substitute for medical diagnosis or treatment.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Irma Darise, Sp.OT, Dr. Irawan Huntoyungo, Sp.OT, and Dr. Chairul Wahjudi, Sp.OT, who have shared their knowledge and expertise in developing this expert system, and Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Negeri Gorontalo which supported the implementation and publication of this research.

REFERENCES

[1] I. Alves *et al.*, "A systematic literature review of the impact and measurement of mobility impairment in rare bone diseases," *Ther. Adv. Musculoskelet. Dis.*, vol. 17, 2025, doi: 10.1177/1759720X251369963.

[2] J. C. P. Butarbutar, P. Basuki, V. Sungono, A. Riantho, and K. Fidiasrianto, "Burden of osteoarthritis in Indonesia: A Global Burden of Disease (GBD) study 2019," *Narra J*, vol. 4, no. 2, pp. 1–10, 2024, doi: 10.52225/narra.v4i2.884.

[3] A. M. Wu *et al.*, "Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019," *Lancet Healthy Longev.*, vol. 2, no. 9, pp. e580–e592, 2021, doi: 10.1016/S2666-7568(21)00172-0.

[4] D. K. Rezha and Z. Ismah, "Analisis Pengaruh Paparan Asap Rokok terhadap Kejadian Osteoporosis pada Wanita Usia 50-59," *Media Kesehatan Masyarakat Indonesia*, vol. 23, no. 2, pp. 134–140, 2024, doi: 10.14710/mkmi.23.2.134-140.

[5] K. RI, "Laporan Nasional Riset Kesehatan Dasar (RISKESDAS)," Lembaga Penerbit Balitbangkes, Jakarta, 2019. [Online]. Available: https://www.badankebijakan.kemkes.go.id/repositori/id/eprint/3514/1/Laporan Riskesdas 2018 Nasional.pdf

[6] B. P. Gorontalo, "PROVINSI GORONTALO DALAM ANGKA 2025," Gorontalo, 2025. [Online]. Available: https://gorontalo.bps.go.id/id/publication/2025/02/28/ec5aa9aee0192d815b59828a/gorontalo-province-in-figures-2025.html

[7] M. of H. of the R. of Indonesia, "Rasio Sp.OT – Spesialis Orthopaedi dan Traumatologi Terhadap Populasi," Indikator SDM Kesehatan. [Online]. Available: https://eplanning-ditjennakes.kemkes.go.id/indikator/user/rasio_populasi_sps/?nakes=5

[8] S. P. B. V. D. Sudhewa, I. G. A. R. Widowati, P. P. Empuadji, N. P. A. Suryaningsih, and I. A. M. P. Sutema, "Self-Medication in the Pandemic Era: Factors to Consider Based on the Theory of Health Belief Model," *PHARMACY: Jurnal Farmasi Indonesia (Pharmaceutical Journal of Indonesia)*, vol. 20, no. 1, p. 59, 2023, doi: 10.30595/pharmacy.v0i0.15071.

[9] F. R. Muharram *et al.*, "Adequacy and Distribution of the Health Workforce in Indonesia," *WHO South. East. Asia J. Public Health*, vol. 13, no. 2, pp. 45–55, Jul. 2024, doi: 10.4103/WHO-SEAJPH.WHO-SEAJPH_28_24.

[10] S. Hartati, *KECERDASAN BUATAN BERBASIS PENGETAHUAN*. Gadjah Mada University Press, 2021.

[11] R. Rahayu, Prastadi Wibawa; Judijanto, Loso; Syaddad, Hasbu Naim; Apriyanto; Adawiyah, *PENGANTAR SISTEM PAKAR (Teori dan Implementasi)*. Yogyakarta: PT Green Pustaka Indonesia, 2025.

[12] D. Reizandi and A. Witanti, "Implementasi Sistem Pakar Diagnosis Penyakit Gigi dan Mulut Berbasis Web dengan Metode Hybrid Chaining," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 12, no. 3, 2025, doi: 10.35957/jatisi.v12i3.12969.

[13] S. B. Atim and M. Y. I. I. Ibrahim, "Rule-Based Expert System Model with Backward Chaining Algorithm for Symptom-Based Skin Disease Diagnosis," *Jurnal Teknologi Dan Open Source*, vol. 8, no. 1, pp. 288–294, 2025, doi: 10.36378/jtos.v8i1.4416.

[14] R. D. Jayanti, B. Rahman, and I. Fitri, "Diagnosa Penyakit Tulang Belakang Menggunakan Metode Forward Chaining dan Certainty Factor," *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, p. 414, 2022, doi: 10.30865/mib.v6i1.3497.

[15] C. Cho, G. Bak, D. Sumpton, B. Richards, and C. Sherrington, "Perspectives of healthcare providers on osteoporosis, falls and fracture risk: a systematic review and thematic synthesis of qualitative studies," *Arch. Osteoporos.*, vol. 19, no. 1, 2024, doi: 10.1007/s11657-024-01446-8.

[16] Aldy Nifratama, Tri Suratno, and Daniel Arsa, "Analisis dan Evaluasi Pengujian pada Penerapan Metode Prototype dalam Software Engineering," *Jurnal Ilmiah Media Sisfo*, vol. 18, no. 1, pp. 128–138, 2024, doi: 10.33998/mediasisfo.2024.18.1.1649.

[17] N. Kania, Y. S. Kusumah, J. A. Dahlan, E. Nurlaelah, F. Gürbüz, and E. Bonyah, "Constructing and providing content validity evidence through the Aiken's V index based on the experts' judgments of the instrument to measure mathematical problem-solving skills," *REID (Research and Evaluation in Education)*, vol. 10, no. 1, pp. 64–79, 2024, doi: 10.21831/reid.v10i1.71032.

[18] P. Leszczyński, P. Lisiński, B. Kwiatkowska, T. Blicharski, J. Drobnik, and K. Pawlak-Buś, "Clinical expert statement on osteoarthritis: diagnosis and

therapeutic choices," *Reumatologia*, vol. 63, no. 2, pp. 104–115, 2025, doi: 10.5114/reum/199980.

[19] O. Suria, "A Statistical Analysis of System Usability Scale (SUS) Evaluations in Online Learning Platform," *Journal of Information Systems and Informatics*, vol. 6, no. 2, pp. 992–1007, 2024, doi: 10.51519/journalisi.v6i2.750.

[20] R. Prihatni, S. P. Hastono, and S. H. Afrizal, "Evaluasi kualitas aplikasi kesehatan menggunakan System Usability Scale Health apps evaluation using System Usability Scale," *Manajemen Informasi Kesehatan*, vol. 9, no. 1, pp. 1–9, 2024.