SymptoTech:Integrative AI System For Symptom Analysis And Customized Health Solutions

Utkarsha Chore[1, a)] and Ehteshamul Haque[1, b)]

Author Affiliations
[1]*Department of Computing Technology,SRM Institute of Science and Technology,KTR,Chennai,India*

Author Emails
[a)]up3406@srmist.edu.in
[b)]eh8302@srmist.edu.in

Abstract. SymptoTech is a comprehensive AI-based healthcare solution designed to offer accurate, personalized diagnostic insights. Specifically developed to address various health conditions with an emphasis on dermatology, SymptoTech integrates image analysis with symptom-based diagnostics to enhance healthcare accessibility. The system employs Convolutional Neural Networks (CNNs) for identifying skin conditions from user-provided images while incorporating Natural Language Processing (NLP) for additional symptom analysis, leading to improved diagnostic precision. Furthermore, SymptoTech features a dietary recommendation engine that provides users with health-specific nutritional guidance. By seamlessly combining data acquisition, AI-driven diagnostics, and personalized health insights, SymptoTech aims to bridge the gap between technology and efficient medical decision-making, ensuring a more accessible and precise healthcare experience.

.

INTRODUCTION

Background

Skin diseases are among the most common health concerns, affecting individuals across all age groups. They can range from non-threatening conditions like eczema and psoriasis to severe diseases such as melanoma, which require immediate medical attention. The identification of skin diseases typically relies on visual assessment by dermatologists, who analyze skin lesions, texture, and color variations. However, the availability of dermatologists is limited, especially in remote areas, leading to delayed diagnosis and treatment.

With the advancement of artificial intelligence (AI) and machine learning (ML), automated skin disease detection has gained significant traction. AI models trained on vast datasets of skin images can help recognize patterns and provide diagnostic suggestions. However, image-based detection alone is not always reliable due to similarities between different skin conditions, variations in skin tones, and external factors such as lighting conditions. Therefore, incorporating patient-reported symptoms alongside image-based diagnosis can significantly enhance accuracy and reliability.

Problem Statement

Despite advancements in AI-driven dermatology, skin disease classification still faces several challenges. Many skin diseases share visual similarities, making it difficult for deep learning models to distinguish between them with high confidence. Variations in skin pigmentation and lighting conditions further impact the accuracy of image-based predictions. Additionally, relying solely on image recognition can lead to misclassification, especially when disease symptoms are not visually prominent. The lack of labeled datasets representing diverse skin tones and conditions also limits the generalizability of AI models. To address these challenges, this study proposes an integrated AI-driven approach that combines image-based detection with symptom-based analysis. By leveraging deep learning for image classification and machine learning models for symptom evaluation, the system aims to improve diagnostic accuracy and provide personalized treatment recommendations..

Importance

The early and accurate detection of skin diseases is crucial in preventing complications and improving patient outcomes. An AI-powered skin disease detection system offers several key benefits, including early detection and timely treatment, enabling patients to seek medical advice before complications arise. It also helps reduce the healthcare burden by assisting dermatologists in filtering out non-critical cases, allowing them to focus on more severe conditions. Additionally, AI enhances accessibility in remote areas by integrating with telemedicine applications, providing healthcare services to individuals with limited access to specialists. Furthermore, by incorporating both major and minor symptoms reported by patients, the system improves diagnostic accuracy, ensuring more precise disease classification.

RELATED WORK

Research in AI-driven dermatology has grown significantly, leveraging deep learning and natural language processing for automated skin disease diagnosis. Existing studies primarily focus on image-based classification, symptom-based diagnosis, or a combination of both. This section reviews prior research and compares different approaches used in the field.

Existing Research

Several studies have explored deep learning for skin disease detection using convolutional neural networks (CNNs). The ISIC Challenge has provided benchmark datasets for melanoma classification, where models like ResNet, VGG, and EfficientNet have achieved dermatologist-level accuracy. These deep learning models have shown significant potential in automating dermatological diagnostics, reducing the dependency on specialist consultations, and providing faster results. However, challenges such as class imbalance, dataset bias, and the need for high-quality labeled data still persist, necessitating continuous advancements in model training and evaluation.

Deep learning has predominantly relied on image-based methods for skin disease classification. Esteva et al. (2017) demonstrated that CNNs trained on over 100,000 skin lesion images could achieve dermatologist-level performance. This landmark study showcased the power of AI in detecting various types of skin conditions with high accuracy. Similarly, Han et al. (2018) conducted a comparative analysis of different CNN architectures and found that ensemble models, which combine multiple deep learning models, consistently outperform individual models by leveraging diverse feature extraction methods. Additionally, Tschandl et al. (2019) explored the effectiveness of transfer learning, revealing that pre-trained models significantly improve accuracy in skin lesion classification. These studies highlight the strengths of image-based deep learning in dermatology but also emphasize the need for complementary approaches to overcome its limitations.

While image-based models focus on visual analysis, symptom-based approaches utilize natural language processing (NLP) to extract critical information from patient descriptions and medical records. NLP models such as BiLSTM and transformers like BERT have been successfully applied to analyze textual descriptions of skin conditions, helping to identify key symptoms that may not be visually detectable. Electronic health records provide valuable data that, when processed with NLP, can enhance diagnostic precision. Additionally, chatbot-based diagnostic systems have been developed to collect structured symptom data using decision trees and symptom checklists. These systems improve accessibility, allowing patients to receive preliminary assessments based on their reported symptoms before seeking professional consultation. Despite their advantages, symptom-based approaches face challenges related to linguistic variations, ambiguity in symptom descriptions, and the need for extensive labeled datasets.

Recent research has focused on integrating image-based CNN models with symptom-based NLP models to create a more comprehensive diagnostic system. For example, Liu et al. (2021) proposed a multi-modal fusion system that combines lesion images with patient-reported symptoms, demonstrating improved classification accuracy compared to single-modality models. Decision-level fusion techniques, where separate models predict independently and their outputs are combined using weighted averaging or attention mechanisms, have also been explored. By leveraging both visual and textual data, hybrid approaches enhance the reliability of AI-driven dermatology by reducing misclassification caused by visual similarities between different skin diseases.

Our research builds upon these existing works by implementing a hybrid AI model that integrates deep learning for image classification with machine learning models for symptom evaluation. This approach aims to improve diagnostic accuracy, enhance model generalization, and provide personalized treatment recommendations. By addressing the limitations of single-modality methods, our integrated system seeks to advance AI-driven dermatology, making it more robust.

Challenges

Developing an AI-driven skin disease detection system presents multiple challenges that need to be addressed for accurate and reliable predictions. One of the primary concerns is data availability and quality. High-quality, labeled dermatological datasets are essential for training deep learning models, but medical data is often scarce, imbalanced, or lacks standardized annotation. Variations in skin tones, lesion types, and imaging conditions further complicate generalization, making it difficult for AI models to perform consistently across different populations.

Another significant challenge is integrating image and text data. Fusing image-based CNN models with symptom-based NLP models requires careful alignment between visual and textual information. Ensuring that symptom descriptions correspond accurately to the given skin images is crucial, as any mismatch can lead to incorrect predictions. Developing a seamless connection between these modalities remains an ongoing challenge in multi-modal AI systems.

Differentiating between similar skin conditions is also a complex task. Many skin diseases, such as eczema, psoriasis, and fungal infections, share overlapping visual and symptomatic characteristics. Deep learning models must learn fine-grained distinctions to avoid misdiagnosis, which often requires large, well-annotated datasets and advanced feature extraction techniques. Without sufficient differentiation, the risk of false positives and negatives increases, affecting the reliability of AI-driven diagnostics.

Generalization across diverse populations is another pressing issue. Most AI models are trained on limited datasets that may not adequately represent diverse skin tones, demographics, or geographical variations. This lack of diversity can result in biased predictions, making the system less effective for underrepresented groups. Ensuring fairness and accuracy across all patient groups is a critical challenge that requires more inclusive dataset collection and bias-mitigation strategies.

Handling subjective symptom inputs adds another layer of complexity. Patients may describe symptoms in different ways, leading to inconsistencies in symptom-based diagnosis. NLP models must be robust enough to understand variations in terminology and descriptions, accounting for linguistic diversity and contextual nuances.

Failure to do so can reduce the accuracy of symptom-based assessments, limiting the effectiveness of hybrid AI systems.

Regulatory and ethical considerations are also paramount in AI-driven medical diagnosis. These systems must comply with healthcare regulations and ethical standards to ensure patient safety and trust. Transparency, interpretability, and accountability in AI predictions are necessary for clinical adoption. Without proper validation and regulatory approval, AI models cannot be reliably integrated into medical practice.

Our research aims to overcome these challenges by implementing a multi-modal fusion approach that integrates image analysis and symptom-based assessment. By combining deep learning for image classification with NLP for symptom evaluation, we aim to develop a more accurate, inclusive, and clinically relevant skin disease detection system. This approach ensures better diagnostic precision, reduces biases, and enhances the overall effectiveness of AI in dermatology.

METHODOLOGY

The methodology of this research focuses on the development of an AI-driven skin disease detection system that integrates both image-based and symptom-based analysis. The system consists of multiple stages, including data collection, preprocessing, model development, symptom classification, and fusion of multimodal inputs for enhanced accuracy.

Dataset

Data Collection

For skin disease detection, the dataset consists of both image-based and text-based data. The image dataset is sourced from publicly available medical repositories such as the ISIC (International Skin Imaging Collaboration) dataset, HAM10000, and dermatology datasets available on Kaggle. These datasets contain labeled images of different skin conditions, including acne, eczema, melanoma, and psoriasis.

Additionally, textual data is collected from clinical reports, patient records, and dermatological symptom databases. This data includes self-reported symptoms such as itching, pain, redness, and scaling, which are crucial for differential diagnosis.

Preprocessing

Before training the model, various preprocessing techniques are applied to both image and text data to ensure consistency and improve model performance. Image preprocessing involves resizing and normalizing images to standardize dimensions and pixel values, applying data augmentation techniques such as rotation, flipping, and contrast adjustments to improve generalization, and using noise reduction filters to eliminate artifacts that might mislead the model. Text preprocessing includes tokenization and lemmatization to break down symptom descriptions into meaningful words and convert them into their base forms, stopword removal to focus on key symptoms, and encoding techniques such as TF-IDF, Word2Vec, or BERT embeddings to convert textual symptoms into numerical representations for model processing.

Model

Architecture

The proposed system employs a hybrid deep learning architecture for accurate skin disease classification. A CNN-based image model, such as ResNet50 or MobileNet, is utilized to classify skin conditions based on visual patterns extracted from lesion images. Simultaneously, an NLP-based text model, such as DistilBERT or BiLSTM, processes symptom descriptions to determine possible diagnoses. A fusion network integrates the outputs of both models through a fully connected layer, ensuring a more comprehensive and reliable prediction.

Feature Extraction

Feature extraction is performed on both modalities. For images, key features such as lesion color, texture, shape, and boundary patterns are extracted to assist in disease classification. For symptoms, the model analyzes symptom frequency, co-occurrence patterns, and severity levels based on past medical data to improve prediction accuracy.

Symptoms

The model categorizes symptoms into major and minor indicators to refine its predictions. Major symptoms, such as lesions for melanoma or rashes for eczema, serve as primary diagnostic markers. Minor symptoms, including mild itching or localized pain, provide additional context to differentiate between visually similar conditions. The system assigns greater weight to major symptoms, as they are more disease-specific, while minor symptoms contribute supplementary information to fine-tune the classification process.

Fusion

To enhance diagnostic accuracy, a multimodal fusion strategy is implemented, combining both image-based and symptom-based probability distributions. The CNN model generates probabilities for different skin diseases based on image analysis, while the NLP model does the same based on symptom data. These probabilities are then fused using weighted averaging or attention mechanisms to produce the final classification.

In cases where image-based and symptom-based predictions conflict significantly, a conflict resolution mechanism is employed. The system utilizes rule-based conflict resolution, setting predefined thresholds to prioritize more reliable predictions. Additionally, ensemble learning techniques are used, where a secondary model is trained to resolve conflicting outputs based on historical cases. Furthermore, a user feedback loop allows users to provide additional inputs to refine ambiguous results, improving the system's adaptability over time.

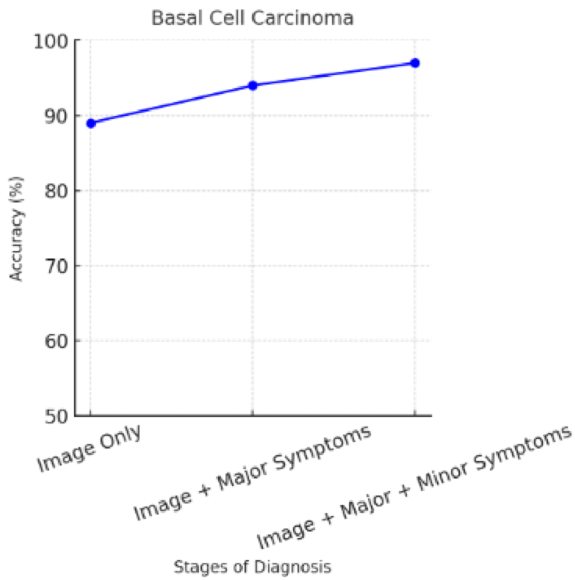
This hybrid approach ensures a more robust, accurate, and user-friendly skin disease detection system by leveraging both visual patterns and patient-reported symptoms in a complementary manner.

RESULTS AND DISCUSSION

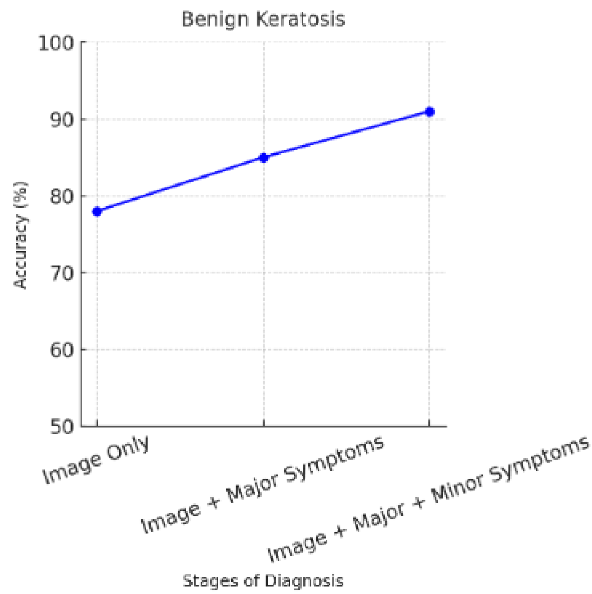
Accuracy

The accuracy of the system is evaluated based on two primary approaches: image-based prediction and symptom-based prediction. The image-based approach utilizes deep learning models trained on labeled skin disease datasets, while the symptom-based approach leverages a structured questionnaire to analyze user-provided symptoms.The hybrid model, which fuses image analysis and symptom-based input, demonstrates improved performance by mitigating false positives and false negatives.

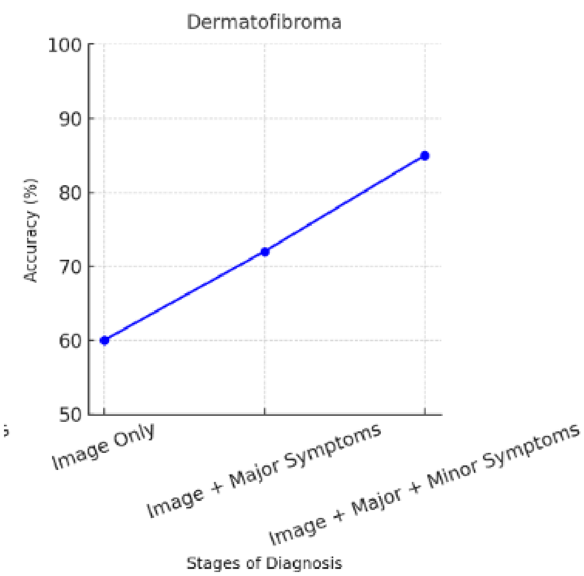
Figures



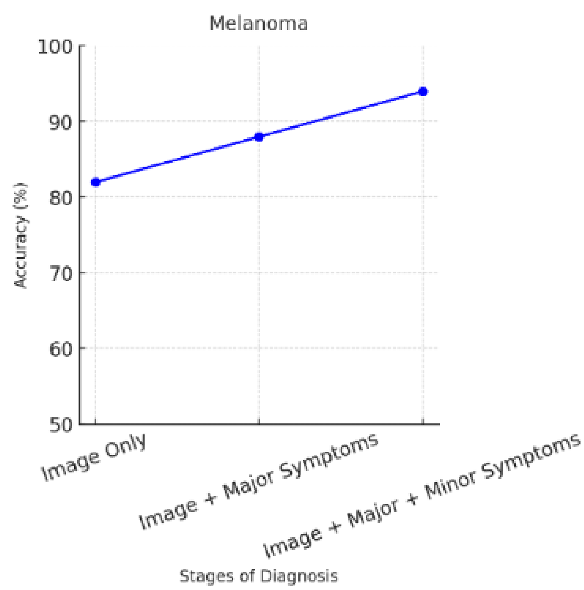
(a)



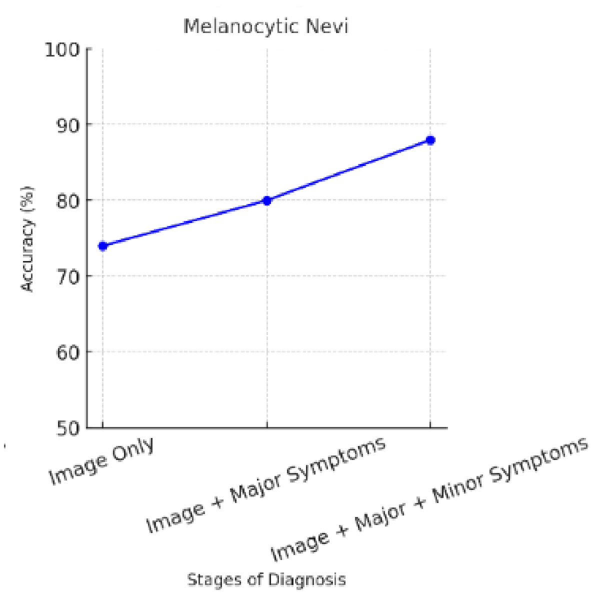
(b)



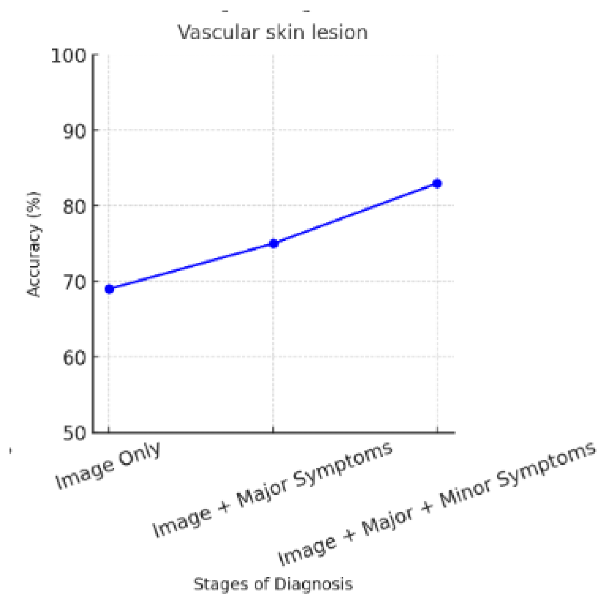
(c)



(d)



(e)



(f)

FIGURE 1. Accuracy progression for skin disease prediction. (a) Accuracy for Basal Cell Carcinoma, (b) Accuracy for Benign Keratosis, (c) Accuracy for Dermatofibroma, (d) Accuracy for Melanoma, (e) Accuracy for Melanocytic Nevi, (f) Accuracy for Vascular Skin Lesion.

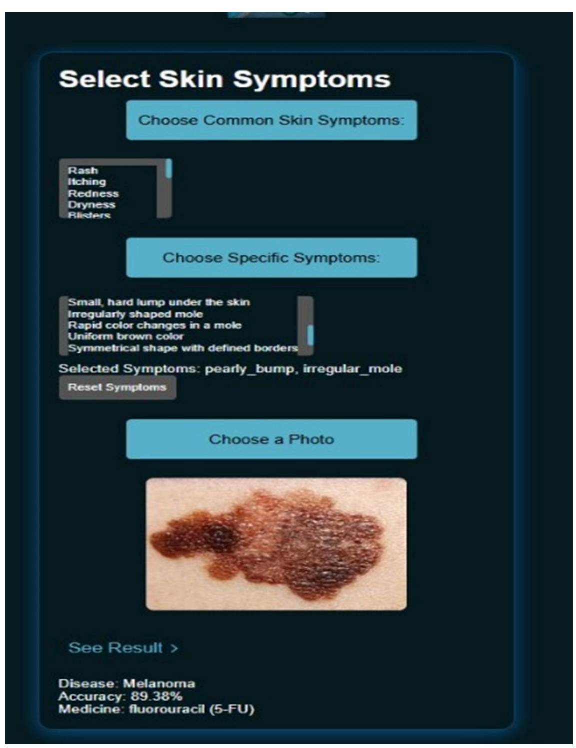


FIGURE 2. Sample output of the prototype for skin disease prediction and treatment recommendation.

Tables

TABLE 1. Accuracy comparison of skin disease prediction using image and symptom-based analysis

Disease Type	Accuracy (Image)	Accuracy (Image + Major Symptom)	Accuracy (Image + Major Symptom + Minor Symptom)
Basal Cell Carcinoma	61.24%	73.39%	86.5%
Benign Keratosis	78.98%	89.21%	92.87%
Dermatofibroma	61.72%	72.4%	85.22%
Melanoma	81.69%	89.36%	93.73%
Melanocytic Nevi	73.65%	80.23%	89.2%
Vascular skin lesion	69.98%	75.54%	83.88%

CONCLUSION

SymptoTech is an advanced AI-driven healthcare solution designed specifically for the accurate diagnosis and personalized analysis of skin diseases. By leveraging Convolutional Neural Networks (CNNs) for image-based disease detection and Natural Language Processing (NLP) for symptom analysis, it ensures a comprehensive approach to dermatological diagnostics. Unlike many existing tools that either focus solely on image-based analysis or symptom-based inputs, SymptoTech integrates both methodologies, improving diagnostic precision and enhancing user experience. Additionally, the system includes a dietary recommendation engine, offering users tailored nutritional guidance based on their diagnosed skin conditions, which helps in managing and preventing further complications. One of the key advantages of SymptoTech is its ability to operate without requiring any physical IoT devices, making it a scalable, cost-effective, and easily accessible solution for users worldwide. With its cloud-based processing, it can be deployed across various platforms, including mobile applications and telemedicine services, ensuring that users can receive instant diagnostic insights from anywhere. Future advancements will focus on enhancing AI accuracy, expanding the database of skin diseases, and integrating with cloud-based electronic health records (EHRs) to provide a more seamless healthcare experience. As digital healthcare continues to evolve, SymptoTech aims to be at the forefront of AI-powered dermatological diagnostics, making early detection, personalized recommendations, and accessible healthcare a reality for millions.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to Dr. M. Mathan Kumar SIE, our project guide, for his valuable guidance and support throughout this research. We also acknowledge the work of Suganiya Murugan, S.R. Srividhya, S. Pradeep Kumar, and B. Rubini, authors of the paper titled "A Machine Learning Approach to Predict Skin Diseases and Treatment Recommendation System", published in the 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), which served as a foundation for our study.

REFERENCES

1. Manish Kumar and Rajiv Kumar, "An intelligent system to diagnosis the skin Disease", vol. 11, no. 19, October 2016.
2. Jerrey Glaister, "Automatic segmentation of skin lesions from dermatological photograph", 2013.
3. K. Indupriya and Dr. G. P. Ramesh Kumar, "A Survey On Skin Texture Analysis For Medical Diagnosis Using Image Processing Techniques", *K. Indupriya Et Al*, vol. 3, no. 5, 2015.
4. Sumithra Ra, Mahamad Suhilb and D.S. Guruc, "Segmentation and Classification of Skin Lesions for Disease Diagnosis", 2015.
5. R. Yasir, M. S. I. Nibir and N. Ahmed, "A skin disease detection system for financially unstable people in developing countries", *Global Science and Technology Journal*, vol. 3, no. 1, pp. 77-93, 2015.
6. V. Kumar, S. Kumar and V. Saboo, "Dermatological Disease Detection Using Image Processing and Machine Learning", *IEEE.*, 2016.

7. Ekta Singhal, "Skin Cancer Detection using Artificial Neural Network", *International Journal of Advanced Research in Computer Science*, vol. 6, no. 1, Jan-Feb 2015.