* Slide 1:

Hi, everyone. My name is Hieu.

Today, I'll be presenting on the topic of **'Non-Invasive Analysis For Health'**.

This presentation is based on the paper **"Contactless Health Monitoring: An Overview of Video-based Techniques Utilising Machine and Deep Learning."**

Let's get started."

* Slide 2: outline

Here is the **Outline** for our discussion today. 1...2...3...4

(Pause)

So, let's get started with our first section: the **Introduce**."

* Slide 4: Introduce

(Point to "Core Technology") First, our **Core Technology** is **iPPG**, I will define this in more detail in just a moment.

(Point to "Our Approach") Second, **Our Approach** to processing the data from this technology involves applying advanced **Machine Learning and Deep Learning** algorithms.

iPPG data can be very 'noisy'. We use ML and DL to **robustly analyze** this noisy data and, from it, accurately extract key health metrics.

(Point to "What is iPPG?") "So, this brings us to the main question: **What is iPPG?**

(Click/Point to the 4 sub-bullets one by one)

- At its core, it's a completely **non-contact technology** that uses a simple camera to measure vital signs.
- It analyzes a video feed to detect **microscopic skin color changes** that are invisible to the naked eye.
- **The Cause** of these tiny changes is simply your **pulse**—or more specifically, the change in blood volume under your skin with each heartbeat.
- And finally, the **Goal** is to use this method to remotely extract signals like **heart rate, breathing rate, and even blood pressure**

* Slide 5: Illustrates

This diagram illustrates the core physics of how iPPG works.

(Point to the left side) "First, **'Ambient Light'** from any shines onto the skin.

(Point to the two reflection arrows) "When it hits the surface, two things happen. A portion of the light returns immediately as **'Specular Reflection'**, which is essentially noise.

However, the light we care about is the **'Diffuse Reflection'**. This light goes through the outer skin layers, it is the **'Epidermis'**—and travels down into the **'Dermis'**.

(Point to the blood vessels) "Down here, it interacts with the **'Capillaries'** and **'Blood Vessels'**. As your heart beats, the blood volume in these vessels changes, which causes a tiny change in how the light is absorbed and reflected back.

(Point to the cameras) "The **'Cameras'** capture all this reflected light—both the noise and the useful signal.

(Point to the right side) "Finally, this video feed is sent to **'Signal Processing'**. This system separates the different color channels (Red, Green, and Blue) and filters out the noise.

The result is the clean, periodic waveform you see on the right: the **'Detected iPPG Signal'**, which directly represents the person's pulse."

* Slide 6: iPPG Processing: ML vs. DL Approaches

(As the slide with Figure 3 appears)

"Here, we compare two primary approaches for processing iPPG signals.

(Point to Figure 3a) Figure (a) shows the **Traditional Machine Learning pipeline**.

- It starts with **'Datasets Collected'** from iPPG
- Crucially, this method requires a dedicated **'Feature Extraction'** step, often using techniques like ICA or PCA, where *we design* what features are important.
- These hand-crafted features are then fed into basic **'ML Algorithms'** like K-means or SVM.
- Finally, these algorithms predict the **'Vital Signs'**. This approach is more manual, requiring explicit feature engineering.

(Point to Figure 3b) In contrast, Figure (b) illustrates the **Deep Learning approach**.

- We monitor specific **'Body Sites'** like the forehead, chest, or palm, defining a **'Region of Interest' (ROI)** based on color or motion.
- The camera captures the raw iPPG waveform directly from these ROIs.
- The key difference is here: the raw iPPG waveform is fed straight into a **'Deep Learning'** model.

- This model automatically learns and extracts the most relevant features on its own, providing a more robust and often more accurate prediction of various **'Vital Signs'**, including more complex ones like Blood Pressure.

* Slide 8: Extract HR based on ML

"(Slide appears) Alright, let's break down the traditional **Machine Learning approach for HR extraction**. This is a classic two-step process.

Step 1 is Region of Interest (ROI) selection. This is the first and most basic step. It simply means we use a tool, like a face detector, to find the specific area on the skin—like the cheeks or forehead—where we will actually measure the signal.

Step 2 is the ML-Based Signal Processing Pipeline. This is the core of the method. As you can see, it's not just one model, but a *series* of algorithms working together.

- We often start with **ICA** to separate the tiny, valuable pulse signal from background noise.
- Then, we might use **Linear Regression (LR)** to find which color channel (Red, Green, or Blue) is giving us the strongest, most reliable signal.
- The **Bayesian approach** is a statistical method. It's used to **synthesize data from two sources**: the skin color changes (which have the pulse) and the head motion (which is noise). By combining these, it can make a more robust prediction of the HR.
- **The Random Forest and K-means** combination is used for advanced noise filtering. First, the Random Forest algorithm is used to distinguish between 'noise' and 'no-noise' groups in the signal. Then, K-means is used to remove the motion noise that was identified.
- Finally, we have **SVM.** This is often used at the end as a quality check. Its job is to **evaluate the reliability** of the final HR measurement. For example, an SVM can be trained to classify a reading as 'correct' or 'incorrect' with very high accuracy—in one study, up to 92.1%.

* Slide 9 : Extract HR based on DL

Now, let's look at the **Deep Learning approach**, which is fundamentally different from the traditional ML pipeline.

The first is the Core Method: End-to-End Learning. Unlike the multi-step ML method, Deep Learning models don't need manual feature extraction. We simply feed the raw video frames or the raw iPPG signal *directly* into the model, and it learns to find the pulse signal on its own.

The second is to look at the Deep Learning Architectures that make this possible. There are three main types:

- First, we have **Hybrid models**, which are very common. These **combine CNN and RNN**. The CNN part analyzes the image frame, while the RNN part analyzes how the pixels change over time.

- Second, **Attention Mechanisms**. Models like DeepPhys and MTTS-CAN use an attention mechanism to automatically focus on the most important regions or frames, which helps to improve the signal.
- And third, **Multi-scale Models**. This is a very popular technique. Models like MSSTNet, GLISNet, and methods using Gaussian pyramids analyze the video at multiple different resolutions. This helps to reduce noise and improve accuracy. Notably, GLISNet has shown clear effectiveness in low-light conditions.

* Slide 11: Extract RR based on ML

"(Slide appears) Now, let's discuss the methods for **extracting Respiratory Rate (RR) based on Machine Learning**.

There are two main methods.

- **Method 1 is using a Thermal Camera.** This technique works by tracking the temperature changes at the nostrils as a person breathes in cool air and exhales warm air. However, this method is not common because thermal cameras are very expensive and impractical.
- Therefore, most research focuses on **Method 2: using a standard RGB Camera** for **Motion Tracking**. This is the more feasible approach. It works by detecting the very subtle physical movements of the chest, abdomen, or even the head as a person breathes.

Now, for this RGB method, several **Specific ML Algorithms** are used to analyze that motion data:

- **Linear Regression (LR)** is often used to estimate the RR from the video frames, and it can also help compensate for motion distortions.
- **Random Forest (RF)** can be trained as a classifier, for instance, to detect conditions like apnea (breathing cessation) using only head motion features.
- A **Binary Decision Tree (DT)** is often used in a multi-modal approach. This means it **combines two different data sources**—like the motion signal and the iPPG skin-color signal—to compute a more accurate RR.
- And finally, **Other Classifiers** like **K-star** and **Rotation Forest** have also been tested to find the most accurate model for RR estimation."

* Slide 12: Extract RR based on DL

"(Slide appears) Next, let's look at extracting Respiratory Rate using **Deep Learning**.

- **Our Focus** here is on using standard, low-cost **RGB cameras**, not expensive thermal ones.
- A **Key Insight** for this method is that analyzing **pixel *motion***—that is, the physical *movement* of the chest—is generally more accurate for breathing than analyzing **pixel *intensity***, or the color changes.
- This has led to several advanced **DL Techniques**:

- **EVM (Eulerian Video Magnification):** This is a powerful technique that *amplifies* the tiny, almost invisible breathing movements of the chest, making them much easier for the model to detect.
- **Optical Flow:** This is often used *with* EVM to precisely *track* the direction and speed of those amplified movements.
- **Spatiotemporal Architectures:** These are complex models, like combinations of **LSTM and 3D-CNNs**. They are 'spatiotemporal' because they learn breathing patterns across both *space* (the chest area) and *time* (the breathing rhythm).
- **Simultaneous Prediction:** This is an efficient approach where a single model uses the iPPG signal to predict **both Heart Rate and Respiratory Rate at the same time**.
- **Hybrid Models:** And finally, this method combines the best of both worlds: using a Deep Learning model as a powerful feature extractor, and then feeding those features into a traditional ML classifier, like an SVM, to get the final result."

* Slide 13: Diagram

"(Slide appears) This diagram provides a great overview, summarizing the different workflows we've been discussing for Respiratory Rate estimation.

First, on the far left, we have the **Inputs**. There are two main sensor types: a standard **RGB Camera** and a specialized **Thermal Camera**.

These inputs lead to two different signal extraction paths:

1. The **RGB Camera** path is used for motion tracking. It performs **Face Tracking** to find the person, and then **Tracks the Chest and Abdomen Movement**. This captures the *physical* motion of breathing.
2. The **Thermal Camera** path is used for temperature tracking. It combines with **Face Detection** (which often uses the RGB camera) to find the nose ROI, and then **Tracks the Nose Temperature Change**—that is, the cool air inhaled versus the warm air exhaled.

Both of these raw signals—whether motion-based or temperature-based—are then processed at the **iPPG-Signal Extraction** stage.

Finally, we have three different **Analysis Models** that can be applied to this signal:

- We can use traditional **ML Techniques** like Random Forest or SVM.
- We can use end-to-end **DL Techniques** like CNN or LSTM.
- Or, we can use **Combined ML & DL** models, such as using a CNN to extract features and a Decision Tree to get the final output.

All three of these analysis pathways lead to the same final goal: to **Estimate the Respiratory Rate**, or RR."

* Slide 15: Body Temperature

Moving on to our final vital sign: **Body Temperature**.

- The **Core Technology** used for non-contact measurement relies on two main inputs: **Thermal Cameras**, which are the most common, or emerging methods using standard **RGB Cameras**.
- The **General Algorithm Workflow** for this is a clear 3-step process:
    1. First, **Face and ROI Detection** is performed, often using a DL model to find the person's face.
    2. Second, **Best-ROI Selection**, where the algorithm isolates the most reliable part of the face for a temperature reading, like the forehead or the inner corner of the eye.
    3. And third, **Temperature Extraction and Classification**, where the temperature is measured and the person is classified, for example, as 'healthy' or 'feverish'.
- This slide lists several **Specific ML/DL Algorithm Examples** used in this workflow:

  - **KNN** and **Logistic Regression** are classic ML models often used for that final **classification** step.
  - **V-TEMP** is a very interesting one. It's a novel method specifically designed for **RGB cameras**, and it works by estimating skin temperature based on the skin's **light reflectance properties**.
  - And we also see **Hybrid DL**. For example, a system might use an **SSD** model—which is a fast object detector—to perform the face detection, and then use a simpler algorithm for the final temperature extraction from that ROI."

* Slide 16: Illustrates

"(Slide appears) This figure shows a more complex, multi-modal system for measuring vital signs, specifically Body Temperature.

First, as seen in **part (b)**, the **physical setup** uses *both* an **RGB camera and a Thermal camera** side-by-side to capture a subject.

Now, let's follow the **workflow in part (c)**:

1. The system gets two simultaneous inputs: an **RGB video** and a **Thermal video**.
2. The **RGB video** is used for **Face Detection and Tracking**, because it's much more accurate for this task.
3. Once the face is found, the system identifies the key **ROIs (Regions of Interest)**. As shown in **part (a)**, these are the **forehead for Body Temperature** and the **nose area for Respiratory Rate**.
4. A critical step is **Image Alignment**. This digitally syncs the two different video feeds, so the forehead in the RGB video matches the *exact* same pixels on the thermal video.
5. Finally, this aligned data is fed into **ML or DL algorithms** which then **Estimate the Body Temperature (BT)**."