

EmotionRate: Wie Filme Emotionen wecken

Eine Streaming-Applikation zur Bereitstellung emotionaler Erfahrungen mit Filmen basierend auf integrierter Emotionserkennung und Langeweile-Analyse

Cristian Gavriliu

M.Sc. HCI, LMU

cristian.gavriliu@campus.lmu.de

Leonie Münster

M.Sc. Media Informatics, LMU

l.muenster@campus.lmu.de

Lea Sigethy

M.Sc. Computer Science, LMU

lea.sigethy@campus.lmu.de

Tudor Teofanescu

M.Sc. Computer Science, LMU

tudor.teofanescu@campus.lmu.de

Affective Computing - Emphatic AI
Ludwig-Maximilians-Universität, München

2. August 2024

Zusammenfassung

Die Entscheidungsfindung bei der Film- oder Serienwahl fällt durch das Überangebot in zahlreichen Online-Streaming-Plattformen oft schwierig, weshalb Möglichkeiten bereitgestellt werden sollten, die bei der Durchsuchung von Filmen und Serien unterstützen. Ein zentraler Aspekt sind hierbei die emotionalen Erfahrungen, die Filme und Serien bieten. Die derzeitig vorwiegende Unterteilung von Filmen und Serien in Genres bietet dabei jedoch keine klare Darstellung der emotionalen Erfahrung und Filme und Serien lassen sich häufig zu mehreren Genres zuordnen. Allerdings ermöglicht der Einbezug von Emotionen eine Abstimmung der individuellen Stimmung mit der emotionalen Erfahrung eines Films oder einer Serie und erleichtert somit die Entscheidungsfindung. Eine vielversprechende Möglichkeit, Emotionen zu erkennen und somit einbezahlen zu können, bietet *Affective Computing*. Durch dessen Anwendung werden intelligente Systeme in die Lage versetzt, menschliche Emotionen zu erkennen, abzuleiten oder zu interpretieren. Die Mehrheit der bestehenden Webseiten und Systeme bietet allerdings keinen Zugang zu Informationen der emotionalen Erfahrung. Daher präsentieren wir *EmotionRate* - Eine Web-Applikation und Online-Streaming-Plattform von Filmen und Serien, die auf der Erkennung von Emotionen und Langeweile basiert, um diese als emotionalen Erfahrung eines Films oder einer Serie bereitzustellen. *EmotionRate* verfolgt dabei das Ziel, die Entscheidungsfindung bei der Film- und Serienauswahl durch die Bereitstellung von Informationen zur emotionalen Erfahrung der einzelnen Filme und Serien zu erleichtern und diese mit der individuellen Stimmung der Nutzenden abzustimmen. Basierend auf den Erkenntnissen, die wir durch unsere Literaturrecherche zu Verwandten Arbeiten gewonnen haben, haben wir eine Web-Applikation implementiert, die in der Lage ist, Emotionen und Langeweile zu erkennen. Dafür wird ein neuronales Netz zur Erkennung der sechs Grundemotionen verwendet. Anschließend führen wir eine Studie durch, um zu evaluieren, ob *EmotionRate* Emotionen korrekt erkennen kann. Darüber hinaus haben wir untersucht, wie das Konzept von den Teilnehmenden wahrgenommen und bewertet wird. Die Ergebnisse zeigen, dass die Genauigkeit bei der Erkennung von Emotionen und Langeweile von Emotion zu Emotion variiert. Dennoch ist ein allgemeiner Trend erkennbar, von dem sich eine emotionale Erfahrung ableiten lässt. Zudem begrüßt die Mehrheit der Teilnehmenden das Konzept von *EmotionRate*. Einige weitere interessante Beobachtungen zeigen schließlich, wie das Konzept in Zukunft erweitert werden könnte.

Inhaltsverzeichnis

1 Motivation	1
2 Verwandte Arbeiten	2
2.1 Erkennung von Emotionen und Langeweile	2
2.1.1 Emotionstheorien	2
2.1.2 Langeweile	2
2.2 Filmauswahl mit Emotionserkennung in Verwandten Arbeiten	3
3 Methodik	3
3.1 Konzept	3
3.1.1 Idee	3
3.1.2 Integrierte Filme	4
3.2 Implementierung	4
3.3 Gesichtserkennung	4
3.4 Emotionserkennung	4
3.5 Langeweileerkennung	5
3.6 Kombination der Komponenten	6
3.7 Web-Applikation	6
3.7.1 Implementierung	6
3.7.2 Datenbankintegration	8
4 Studie	8
4.1 Studiendesign	8
4.1.1 Teilnehmende	8
4.1.2 Studienablauf	9
4.2 Evaluation	9
4.2.1 Quantitative Daten	9
4.2.2 Qualitative Daten	10
4.2.3 System Usability Score	10
5 Diskussion	10
5.1 Zukünftige Arbeiten	12
6 Fazit	12

1 Motivation

Mit fortschreitender Digitalisierung hat sich der Konsum von Filmen und Serien erheblich gewandelt und findet nicht länger nur in Kinosälen oder vor dem Fernseher statt. Stattdessen sind immer mehr Filme und Serien über Online-Streaming-Plattformen verfügbar, die über digitale Endgeräte, wie Computer und Smartphones, zugänglich sind. In diesem Kontext wurde das Streaming von Filmen und Serien flexibler und individueller, da digitale Endgeräte bisherige Möglichkeiten des Streamings erweitern, sowie Filme und Serien jederzeit und überall abgerufen werden können. Infolgedessen haben viele Applikationen für Online-Streaming-Plattformen Einzug auf Computer und Smartphone erhalten.

Jedoch stellen diese reichhaltigen Umgebungen, die große Anzahl an verfügbaren Filmen und Serien, sowie deren rasch wachsendes Angebot eine erhebliche Herausforderung für die Entscheidungsfindung bei der Film- oder Serienauswahl dar [8]. Um diese zu bewältigen, sind daher neue Methoden erforderlich, die das Finden passender Filme und Serien unterstützen [14, 37]. Ein zentraler Aspekt hierbei sind die emotionalen Erfahrungen der Filme und Serien [8].

Filme und Serien bieten komplexe emotionale Erfahrungen und lösen eine Vielfalt an Emotionen aus, weshalb sie daher oft als *Emotionsmaschinen* bezeichnet werden [8, 33]. Darüber hinaus stehen Emotionen in direktem Zusammenhang mit der Konsumerfahrung von Filmen und Serien und stellen dabei einen entscheidenden Faktor dar, der bei der Entscheidungsfindung berücksichtigt werden sollte [37]. Durch Emotionen ausgelöste Reaktionen, sogenannte Affekte, sind nicht nur ein wesentlicher Grund für den Erfolg eines Films, sondern auch einer der wichtigsten Anreize für das Ansehen von Filmen und Serien [15, 33]. Daher sollten emotionale Erfahrungen nicht als zusätzliches Merkmal, sondern als zentraler Bestandteil der Film- oder Serienauswahl betrachtet werden [8, 33].

Um die Entscheidungsfindung zu erleichtern, werden Filme und Serien derzeit vorwiegend in Kategorien nach emotionaler Bedeutung, sogenannten Genres, klassifiziert, was jedoch nicht immer die tatsächliche emotionale Erfahrung widerspiegelt [8]. Die Problematik bei der Unterteilung von Filmen und Serien in Genres liegt vor allem darin begründet, dass Filme und Serien Erlebnisgüter sind, weshalb ihr Wert vor dem Konsum schwer zu beurteilen ist [25]. Genres bieten keine klare Darstellung der emotionalen Erfahrung und zudem lassen sich Filme und Serien häufig zu mehreren Genres zuordnen [23]. Lanir et al. [8] betont deshalb die Notwendigkeit, Möglichkeiten bereitzustellen, Filme und Serien entsprechend ihrer emotionalen Erfahrung zu erkunden. Der Einbezug von Emotionen ermöglicht dabei eine Abstimmung der individuellen Stimmung mit der emotionalen Erfahrung eines

Films oder einer Serie und erleichtert somit den Entscheidungsfindungsprozess. Daher sollten Emotionen für Empfehlungen bei der Film- oder Serienauswahl genutzt werden. [8, 18, 29, 36, 39].

Eine vielversprechende Möglichkeit, Emotionen zu erkennen und somit einbeziehen zu können, bietet *Affective Computing*. Dabei handelt es sich um ein zunehmend wichtiges, interdisziplinäres Forschungsfeld, das darauf abzielt, menschliche Emotionen mit Hilfe intelligenter Systeme zu erkennen, abzuleiten oder zu interpretieren [34]. *Affective Computing* kombiniert Elemente aus der Psychologie, Kognitionswissenschaft, Physiologie und Informatik und bietet somit die Möglichkeit, Emotionen zu erkennen, um die emotionale Erfahrung von Filmen und Serien zugänglich zu machen. Die Kombination der Vorteile des Online-Streamings mit *Affective Computing* könnte daher ganz neue Möglichkeiten und Nutzen im Vergleich zur herkömmlichen Entscheidungsfindung bei der Film- und Serienauswahl bieten.

Die Mehrheit der bestehenden Webseiten und Systeme bietet zudem keinen Zugang zu Informationen, die in emotionaler Form vermittelt werden [17]. Darüber hinaus berücksichtigen bisher nur wenige Systeme die affektiven Aspekte von Filmen und Serien [8]. Aufgrund der starken emotionalen Wirkung von Filmen können Visualisierungstechniken, die Emotionen einbeziehen, dabei helfen, die Komplexität der Filminformationen zu bewältigen [8, 33].

Daher präsentieren wir *EmotionRate* - eine Webapplikation, die auf der Erkennung von Emotionen und Langeweile basiert, um somit Informationen zur emotionalen Erfahrung von Filmen und Serien in einer Streaming-Applikation bereitzustellen. Diese Idee entspricht somit der Forderung der Bereitstellung von Möglichkeiten, Filme und Serien basierend auf ihrer emotionalen Erfahrung zu erkunden [8]. *EmotionRate* verfolgt dabei das Ziel, die Entscheidungsfindung bei der Auswahl von Filmen oder Serien durch die Bereitstellung von Informationen zur emotionalen Erfahrung zu erleichtern und eine Möglichkeit bereitzustellen, die Auswahl mit der individuellen Stimmung abstimmen zu können.

Um dies zu erreichen, konzentriert sich diese Arbeit auf die Erkennung von Grundemotionen sowie von Langeweile. Es wurden zunächst die theoretischen Grundlagen für die Erkennung von Emotionen und Langeweile untersucht und bestehende Ansätze analysiert. Anschließend wurde im Rahmen des Projekts eine Web-Applikation entwickelt, die die Erkennung von Emotionen sowie die Analyse der Langeweile mittels eines *Convolutional Neural Networks* und *Facial Landmarks* implementiert. Nachdem die Implementierung abgeschlossen war, folgte eine Studie, um zu evaluieren, ob *EmotionRate* die gewünschten Ergebnisse liefert. Im Zentrum der Untersuchung stand, ob die erkannten Emotionen und die Langeweile mit der Selbst-

einschätzung der Teilnehmenden übereinstimmt. Zum anderen wurde untersucht, ob die Nutzenden das Konzept der automatischen Erkennung von Emotionen und Langeweile sowie die Darstellung dieser Informationen als emotionale Filmerfahrung positiv bewerten. Daraufhin wurden die Ergebnisse der Studie analysiert und ausgewertet. Die Ergebnisse zeigen, dass die Genauigkeit bei der Erkennung von Emotionen und Langeweile deutlich von Emotion zu Emotion variiert und Genauigkeiten zwischen 18,18% für die Emotion *Abscheu* sowie 97,33% für die Emotion *Angst* aufweist. Dennoch ist ein allgemeiner Trend erkennbar, von dem sich eine genaue emotionale Erfahrung ableiten lässt. Zudem begrüßt die Mehrheit der Teilnehmenden das Konzept von *EmotionRate*. *EmotionRate* erhielt ebenfalls eine gute Bewertung der Usability. Abschließend bietet dieses Paper eine Diskussion, die weitere Beobachtungen und Ergebnisse evaluiert und einen Ausblick auf die Zukunft bereitstellt.

2 Verwandte Arbeiten

2.1 Erkennung von Emotionen und Langeweile

Um sicherzustellen, dass die Erkennung der Emotionen in *EmotionRate* fundiert und verlässlich ist, haben wir uns intensiv mit bedeutenden Emotionstheorien beschäftigt. Darauf folgte die Auseinandersetzung mit der Analyse und Erkennung von Langeweile. Das folgende Kapitel bietet dabei einen Überblick über die wichtigsten Erkenntnisse.

2.1.1 Emotionstheorien

Obwohl Emotionen in einer Vielzahl an Forschungsberichen von großer Bedeutung sind und deren Erfassung ein häufig angestrebtes Ziel darstellt, ist die Forschung zu Emotionen dennoch äußerst komplex [8, 28]. So gibt es keine allgemeingültige Definition von Emotionen, da sie ein sehr persönliches inneres Erleben darstellen und oft in Kombination mit weiteren Emotionen auftreten [8, 28]. Im Allgemeinen können Emotionen im Alltag als ein Gefühl oder ein innerer Zustand bezeichnet werden, die häufig durch spezifische Reize ausgelöst werden [8].

Um die Dynamik von Emotionen besser zu verstehen und zu erklären, wurden zahlreiche Emotionstheorien entwickelt. Dabei lassen sich die Emotionstheorien hauptsächlich in zwei große Ansätze unterteilen: dimensionale und diskrete Emotionstheorien [8].

Dimensionale Emotionstheorien modellieren Emotionen anhand mehrerer kontinuierlicher Dimensionen [8]. Das zweidimensionale Modell von Barrett und Russel [30] beispielsweise kategorisiert Emotionen entlang der beiden Achsen Valenz (Freude oder Unbehagen)

und Erregung (hoch oder niedrig). Ein emotionaler Affekt kann somit in diesem zweidimensionalen Raum eingeordnet werden.

Im Gegensatz dazu gehen diskrete Emotionstheorien davon aus, dass alle Menschen über eine Reihe von kulturgebundenen, universellen Grundemotionen verfügen [8]. Eine der bekanntesten diskreten Emotionstheorien stammt von Ekman et al. [10, 11]. Diese Theorie betont, dass Emotionen nicht als einzelne affektive Zustände existieren, sondern als eine Familie verwandter emotionaler Zustände, und identifiziert sechs universelle Grundemotionen. Neben der *neutralen* Emotion sind diese *Wut*, *Abscheu*, *Angst*, *Traurigkeit*, *Freude* und *Überraschung*.

Obwohl sowohl diskrete als auch dimensionale Modelle in der Forschung zur Visualisierung von Emotionen verwendet werden, wird das diskrete Modell häufiger angewandt [24, 8]. In diesem Paper nutzen wir Ekman et al. [10, 11] Emotionstheorie und implementieren eine Applikation, die die Emotionen der Zuschauenden von Filmen oder Serien gemäß Ekman et al. [10, 11] Grundemotionen analysiert.

2.1.2 Langeweile

Neben den Grundemotionen nach Ekman et al. [10, 11] erkennt *EmotionRate* ebenfalls Langeweile bei Zuschauenden von Filmen oder Serien. Langeweile kann sich durch verschiedene körperliche Ausdrucksformen zeigen, die sowohl die Körperhaltung, die Gestik, als auch die Mimik umfassen [12].

Da es sich bei der Langeweile um einen komplexen emotionalen Zustand handelt, ist ihre Definition schwierig [9]. Es gibt jedoch Merkmale der Langeweile, die leichter zu spezifizieren und daher leichter zu erkennen sind. Eines dieser Merkmale ist Müdigkeit [20, 9, 12].

Es gibt bereits zahlreiche Forschungsarbeiten zur Müdigkeitserkennung, insbesondere im medizinischen Bereich und zur Erhöhung der Sicherheit im Straßenverkehr. Müdigkeit kann dabei durch die Öffnung der Augenlider und der Lippen, die Blickrichtung sowie bestimmte Körperhaltungen, wie das Stützen des Kopfs mit der Hand oder das Verdecken von Teilen des Gesichts, erkannt werden [12, 32, 20]. Dabei liegt die Augenöffnung jeweils unter einem bestimmten Grenzwert, da die Augen geschlossen oder nur leicht geöffnet sind [35, 16, 12]. Die Mundöffnung liegt dagegen über einem bestimmten Grenzwert, da der Mund in Form eines Gähnens weit geöffnet ist [35, 12].

In ihren Studien erkennen Zhang et al. [38] sowie Zhuang et al. [40] Müdigkeit beim Autofahren anhand der Weite der Augenöffnung. Zhuang et al. [40] bestimmen dabei zunächst die Augen mithilfe von *Facial Landmarks* der *Dlib*-Bibliothek und berechnen anschließend den Lidverschluss über den PERCLOS-Wert [2], einen in der

Müdigkeitserkennung gebräuchlichen Parameter.

Thulasimani et al. [35] entwickelten einen Prototypen eines Müdigkeitserkennungssystems beim Autofahren, der geschlossene Augen, Gähnen sowie die Kopfneigung erkennt und Warnungen ausgibt, wenn Müdigkeit festgestellt wird. Für die Erkennung der Gesichtspunkte verwendeten sie ebenso die *Facial Landmarks* der *Dlib* Bibliothek. Ob ein Auge offen oder geschlossen ist, wird durch das *Eye-Aspect-Ratio* bestimmt, das sich aus der Division der Höhe des Auges durch die Breite des Auges ergibt. In Bezug auf das Gähnen wird das *Mouth-Aspect-Ratio* berechnet, das sich aus der Division der Höhe der Lippen durch die Breite der Lippen ergibt.

Muñoz et al. [12] nutzen ebenfalls die Erkennung Augen- und Lippenöffnung sowie der Blickrichtung, um Langeweile von Schülern und Schülerinnen zu erkennen und somit Lehrende zu informieren, sodass diese ihre Strategie anpassen können, um die Schüler und Schülerinnen wieder zu fesseln. Darüber hinaus entwickelten Ochocki und Sawicki [26] einen Algorithmus zur Müdigkeitserkennung durch Gähnen, indem sie die maximalen Höhen des Mundes über die Zeit aufzeichneten. Daraus ging hervor, dass die maximalen Höhen beim Gähnen größer sind als beispielsweise beim Lachen.

Müdigkeit kann sich auch in vielen anderen Körperteilen äußern [12]. Da jedoch für die Implementierung dieses Projekts ausschließlich Gesichtsdaten, die von der Frontkamera bereitgestellt werden, erfasst werden und insbesondere die Mimik von besonderem Interesse ist, werden diese Ansätze hier nicht weiter aufgeführt. Der Fokus in unserem Projekt liegt auf der Erkennung von Langeweile durch die Mimik.

2.2 Filmauswahl mit Emotionserkennung in Verwandten Arbeiten

Mehrere Forschungsprojekte haben sich bereits mit der Integration der Erkennung von Emotionen zur Unterstützung der Entscheidungsfindung bei der Film- oder Serienauswahl beschäftigt. Lanir et al. [8] entwickelten beispielsweise eine *Movie Emotion Map*, ein interaktives Tool zur Erkundung von Filmen basierend auf ihrer emotionalen Signatur. Dabei werden Filme anhand ihrer emotionalen Erfahrung in die *Movie Emotion Map* eingeordnet und positioniert, wobei ebenfalls detailliertere Informationen zur emotionalen Erfahrung eines spezifischen Films integriert sind. Allerdings basiert die emotionale Erfahrung eines Films hierbei nicht auf der Emotionserkennung bei Zuschauenden, sondern auf der Analyse und Erkennung von Emotionen in den Filmrezensionen.

Shamoi et al. [18] entwickelten eine Methodik für die Filmauswahl in Gruppen, um den Konsens, die Zufriedenheit, sowie die Entscheidungssicherheit bei der Filmauswahl in Gruppen durch die emotionalen Erfahrungen von Filmen zu erleichtern. Dafür stellten

sie Empfehlungen von Filmen bereit, die am ähnlichsten zu den emotionalen Erfahrungen der Lieblingsfilme aller Gruppenmitglieder sind. Jedoch wurde auch hier nicht die Emotionserkennung bei Zuschauenden für die emotionale Erfahrung von Filmen genutzt, sondern die emotionale Erfahrung anhand von Analysen des Filmplakats, der Filmbeschreibung sowie des Filmsoundtracks bestimmt.

In beiden Studien zeigten sich dabei deutlich positive Erfahrungen bei der Suche eines Films anhand seiner Emotionen.

3 Methodik

3.1 Konzept

3.1.1 Idee

Die Idee des Konzepts von *EmotionRate* umfasst die Bereitstellung von Informationen über die emotionale Erfahrung von Filmen und Serien basierend auf der Emotions- und Langeweileerkennung bei Zuschauenden. Sie zielt darauf ab, die Entscheidungsfindung bei der Film- und Serienauswahl zu unterstützen und zu erleichtern und bietet Nutzenden die Möglichkeit, die Auswahl eines Films oder einer Serie entsprechend der individuellen Stimmung anzupassen. Die Entscheidungsfindung soll effizienter und individueller gestaltbar sein, sodass Nutzende nicht länger mit Unsicherheiten und Überraschungen bezüglich der ausgelösten Emotionen bei der Auswahl zu kämpfen haben. Da diese Unsicherheiten ein häufiges Problem bei traditionellen Streaming-Plattformen von Filmen darstellen, zielt *EmotionRate* darauf ab, diese Herausforderung durch die Integration einer Emotions- und Langeweileerkennung zu adressieren.

Nutzende haben dabei die Möglichkeit, die Aufzeichnung ihrer individuellen Emotionen, ein sogenanntes *Emotion Recording*, während des Anschauens eines Films oder einer Serie zu starten. Die erkannten Emotionen und Anzeichen von Langeweile werden anschließend als zusätzliche Informationsdetails zur emotionalen Erfahrung des Films oder der Serie hinzugefügt. Auf diese Weise bietet sich eine weitere Entscheidungsgrundlage für die Film- und Serienauswahl, die sich an der individuellen Stimmung orientiert und die Zufriedenheit der Auswahl erhöhen soll. Die Zielgruppe umfasst dabei alle, die Online-Streaming-Applikationen nutzen wollen.

Für die konkrete Realisierung der Ziele unseres Konzepts integriert *EmotionRate* die Grundemotionen nach Ekman et al. [10, 11]. Darüber hinaus wird Langeweile durch ihr spezifisches Merkmal der Müdigkeit anhand der Augen- und Mundöffnung der Zuschauenden erkannt. Hierbei ist der Grad der Öffnung dieser Komponenten entscheidend: So wird ein Nutzender als gelangweilt betrachtet, sofern die Augen nur gering



Abbildung 1: Grundlegende Schritte der Implementierung von *EmotionRate*

geöffnet sind und ein weit geöffneter Mund als Gähnen interpretiert wird.

Diese Idee wurde in der Web-Applikation *EmotionRate* umgesetzt, die Filme verschiedener Genres enthält und in der Lage ist, Emotionen und Langeweile in Echtzeit zu erkennen. Um sicherzustellen, dass die Idee fundiert und für zukünftige Arbeiten nutzbar ist, wurde eine kleine Studie durchgeführt. Diese Studie zielt darauf ab, festzustellen, ob die Emotionserkennung mit der Selbsteinschätzung der Teilnehmenden übereinstimmt. Schließlich wurde zusätzlich Feedback zu dem Konzept, den Funktionalitäten sowie der Benutzerfreundlichkeit von *EmotionRate* eingeholt.

3.1.2 Integrierte Filme

Die Filme, die in *EmotionRate* integriert sind, bilden verschiedene Genres ab und variieren in der Länge. Durch dieses breite Spektrum wurde sichergestellt, dass die vielfältigen Interessen der Nutzenden in *EmotionRate* berücksichtigt werden und die Web-Applikation eine breite Zielgruppe anspricht.

3.2 Implementierung

Für die Implementierung wurde eine *Visual Studio Code*-Umgebung mit Python in der Version 3.12.4 genutzt. Der gesamte Code ist in einem GitLab-Repository einsehbar [31]. Dieses Kapitel bietet zunächst einen Überblick über die Gesichtserkennung. Anschließend folgen die Emotions- und Langeweileerkennung, sowie deren Kombination. Schließlich wird ebenfalls ein Überblick über die Web-Applikation bereitgestellt. Abbildung 1 veranschaulicht dabei die einzelnen Schritte der Implementierung, die im Folgenden thematisiert werden.

3.3 Gesichtserkennung

Sofern das *EmotionRecording* in der Web-Applikation gestartet wurde, wird die Kamera des entsprechenden digitalen Endgeräts verwendet, die die Gesichtserkennung der Nutzenden ermöglicht.

Für die Erfassung des Gesichts auf dem Kamerabild und die Anzeige von Textinhalten über dem Kamerabild wurden die Bibliotheken *OpenCV* [5], eine Open-Source-Bibliothek für Computer-Vision- und Machine-Learning-Software, und *Dlib* [19] genutzt.

Als visuelle Rückmeldung für die Nutzenden wird das Gesicht durch ein Rechteck, sowie die Augen

und Lippen durch Linien umrahmt. Abbildung 2 visualisiert die Echtzeit-Erkennung von Emotionen und Langeweile in einem Video-Feed sowie die Umrahmung der *Facial Landmarks*. Die Umrahmung der Augen und Lippen kann dabei die flüssige Leistung der Live-Kamera beeinträchtigen, weshalb diese standardmäßig deaktiviert ist. Nutzende haben jedoch die Möglichkeit, diese Funktion über eine Checkbox in der Benutzeroberfläche zu aktivieren, um die visuelle Rückmeldung zu gewährleisten. Ist die Checkbox nicht aktiviert, erfolgt dennoch jederzeit die Erkennung der Augen und Lippen im Hintergrund, lediglich die Umrahmungen werden nicht angezeigt.

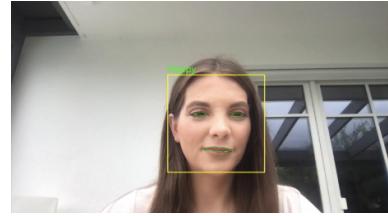


Abbildung 2: Echtzeit Erkennung der Emotion *Freude* sowie Umrahmung der *Facial Landmarks* der Augen und Lippen

3.4 Emotionserkennung

Um zusätzlich die Emotion der Zuschauenden zu erkennen und als Textausgabe im Kamerabild zurückzugeben zu können, wurde im Hintergrund ein *Machine Learning Model* implementiert, das auf den Bibliotheken *Tensorflow* [13] (Version 2.16.1), ein Open-Source-Framework für maschinelles Lernen und künstliche Intelligenz von *Google*, und *Keras* [7] (Version 3.3.3), eine in *Tensorflow* integrierte Open-Source-Bibliothek für neuronale Netze, basiert. Abbildung 2 zeigt dabei ein Beispiel für die Echtzeit-Erkennung von Emotionen und Langeweile in einem Video-Feed sowie die Umrahmung der *Facial Landmarks*.

Jeder Layer des vierstufigen *Convolutional Neural Networks* erkennt verschiedene komplexe Merkmale eines Gesichts, sodass das Modell das Gesicht einer der Grundemotionen nach Ekman et al. [10, 11] zuordnen kann. Das Modell empfängt Bilder mit einer Größe von 48x48 Pixel als Input. Der erste Layer besitzt eine Kernel-Größe von 3x3 und 64 Filter. Der zweite Layer hat eine Kernel-Größe von 5x5 und 128 Filter. Der dritte und der vierte Layer haben jeweils eine Kernel-Größe von 3x3 und 512 Filter. Die Layer sind unterstützt durch eine Batch-Normalisierung, der Aktivierungsfunktion *ReLU*, einer Max-Poolingfunktion mit der Pool-Größe 2x2 und einem Dropout von 0.25. Anschließend wird das geflattete Modell mit zwei Dense-Layern komplett verbunden. Die Dense-Layer sind gefolgt von einer Batch-Normalisierung, der Aktivierungsfunktion *ReLU* und einem Dropout von 0.25 zur Parameterre-

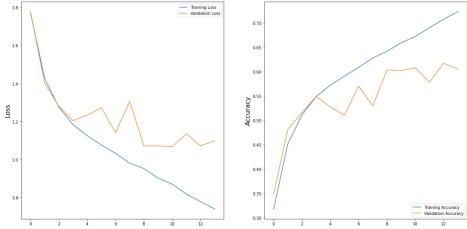


Abbildung 3: Accuracy und Loss während des Trainings des Modells

duktion. Als Optimierungsfunktion wurde *Adam* mit einer Lernrate von 0.0001 genutzt. Der *Loss* wird durch die *Categorical Crossentropy* berechnet.

Das Modell wurde selbst trainiert auf dem *Face expression recognition* Datensatz von Jonathan Oheix aus dem Jahr 2018 [27]. Der Datensatz ist eine neuere Version des *Google*-Datensatzes *FER 2013* und enthält 35.887 Bilder der Grundemotionen, die in einen Trainingsdatensatz von 28.821 Bildern und einen Testdatensatz von 7.066 Bildern aufgeteilt wurden. Für das Training sind dies 3.993 Bilder für die Emotion *Wut*, 436 für *Abscheu*, 4.103 für *Angst*, 4.938 für *Traurigkeit*, 4.982 für *Neutral*, 7.164 für *Freude* und 3.205 für *Überraschung*. Die verbleibenden Bilder werden für den Test verwendet: 960 Bilder für die Emotion *Wut*, 111 für *Abscheu*, 1.018 für *Angst*, 1.139 für *Traurigkeit*, 1.216 für *Neutral*, 1.825 für *Freude* und 797 für *Überraschung*. Die Bilder sind alle in Graustufen, haben eine Größe von 48x48 Pixel und sind auf das Gesicht zugeschnitten.

Unser Modell wurde dabei auf 48 Epochen trainiert. Der Modellverlauf in Bezug auf *Accuracy* und *Loss* über alle Epochen ist in Abbildung 3 veranschaulicht. Wie Abbildung 3 zeigt, hat sich die *Validation Accuracy*, sowie der *Validation Loss*, bis Epoche 14 nicht mehr weiter verbessert, sodass ein *EarlyStopping* das Training nach Epoche 14 beendet hat. Das *EarlyStopping* verhindert dabei, dass das Modell auf den Trainingsdaten auswendiglernt und eine verfälschte Performance aufweist. Das Modell resultierte in einer *Validation Accuracy* von etwa 61% bei einem *Validation Loss* von etwa 1,1 auf einer Skala von 0 bis 2. Das Jupyter-Notebook (*FER.ipynb*) [3] und das Modell (*model.h5*) sind im GitLab-Repository [31] unter /Model zu finden.

3.5 Langeweileerkennung

Im Rahmen des Projekts wird Langeweile anhand von Symptomen der Müdigkeit erkannt, wie in Kapitel 2.1.2 erläutert. Müdigkeit wird dabei hauptsächlich durch eine geringe Augenöffnung und eine große Mundöffnung beim Gähnen identifiziert. Wir nutzen die *Facial Landmarks* von *Dlib*, um sowohl die Augenöffnung als auch die Mundöffnung zu bestimmen.

Öffnung der Augen Um den Öffnungsgrad der Augen zu bestimmen, werden sowohl für das linke als auch für das rechte Auge jeweils sechs *Facial Landmarks* mit ihrer zugehörigen x- und y-Koordinaten bestimmt. Diese *Facial Landmarks*, p1, p2, p3, p4, p5 und p6, sind dabei symmetrisch entlang des Umrisses des Auges angeordnet. Schließlich wird aus diesen Facial Landmarks sowohl für das linke als auch für das rechte Auge der *Eye-Aspect-Ratio* berechnet, der sich aus der folgenden Formel ergibt:

$$EyeAspectRatio = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2 \|p_1 - p_4\|} \quad (1)$$

Sofern der Eye Aspect Ratio unter einen bestimmten minimalen Grenzwert fällt, der in unserer Implementierung bei 0,26 liegt, können wir annehmen, dass das Auge müde, also gering geöffnet oder sogar geschlossen ist.

Öffnung des Mundes Um den Öffnungsgrad des Mundes zu bestimmen, werden insgesamt 20 *Facial Landmarks* für die Ober- und Unterlippe mit ihren zugehörigen x- und y-Koordinaten bestimmt. Anschließend wird jeweils die Mitte der Oberlippe sowie die Mitte der Unterlippe als Mittelwert der y-Koordinaten der entsprechenden *Facial Landmarks* ermittelt. Um nun ein Gähnen zu erkennen, wird die Distanz zwischen der Mitte der Oberlippe und der Mitte der Unterlippe berechnet:

$$lip_distance = upperLip_center - lowerLip_center \quad (2)$$

Sofern die Lippendistanz einen bestimmten maximalen Grenzwert überschreitet, der in unserer Implementierung bei 70 liegt, wird ein weit geöffneter Mund und somit ein potenzielles Gähnen erkannt. Um ein tatsächliches Gähnen von zufälligen Bewegungen zu unterscheiden, wird zusätzlich die Dauer der Lippendistanz überwacht. Ein Gähnen wird erst dann als erkannt betrachtet, wenn die Lippendistanz für eine bestimmte Anzahl von aufeinanderfolgenden Frames den Grenzwert überschreitet, wobei diese Anzahl von Frames in unserer Implementierung bei drei liegt.

3.6 Kombination der Komponenten

Nachdem die Erkennung der Grundemotionen und der Langeweile im Detail betrachtet wurden, präsentiert das folgende Kapitel den kombinierten Ablauf dieser beiden Erkennungen. Abbildung 4 veranschaulicht dabei die Struktur unseres entwickelten Konzepts.

Sobald eine Emotion erkannt wird, überprüfen wir zunächst, um welche Emotion es sich handelt. Handelt es sich nicht um die Emotion *Neutral*, wird sie der emotionalen Erfahrung des Films oder der Serie als Informationsdetail hinzugefügt. Die Emotion *Neutral* stellt dabei einen Sonderfall dar. Da diese Emotion bei diesem Projekt keinen Mehrwert als Informationsdetail zur emotionalen Filmerfahrung liefert, wird sie ignoriert, insofern sie erkannt wird.

Bei der Erkennung von Müdigkeit und somit auch Langeweile standen wir vor der Herausforderung, andere Emotionen, die ähnliche Gesichtsausdrücke wie Müdigkeit und Langeweile aufweisen, nicht fälschlicherweise als Langeweile zu erkennen. Das System muss in der Lage sein, diese Hinweise zu interpretieren und sinnvoll kombinieren zu können, um präzise vorherzusagen, ob es sich um Müdigkeit und somit auch Langeweile handelt.

In diesem Kontext überprüfen wir kontinuierlich den Öffnungsgrad der Augen und des Mundes der Zuschauenden. Wird ein Gähnen erkannt und sind zudem die Augen müde, also nur gering geöffnet oder geschlossen, wird Müdigkeit und somit auch Langeweile erkannt. Wird jedoch ein Gähnen erkannt, während die Augen nicht müde, also normal oder weit geöffnet sind, wird diese Information ignoriert.

Indem wir Müdigkeit und Langeweile nur als Kombination von Gähnen und müden Augen erkennen, können wir sicherstellen, dass spezifische Emotionen wie Freude oder Überraschung, die häufig ebenfalls durch eine große Mundöffnung gekennzeichnet sind, oder Traurigkeit, die oft durch gering geöffnete Augen gekennzeichnet ist, nicht fälschlicherweise als Langeweile interpretiert werden, da sie jeweils das andere Merkmal in der Regel nicht erfüllen.

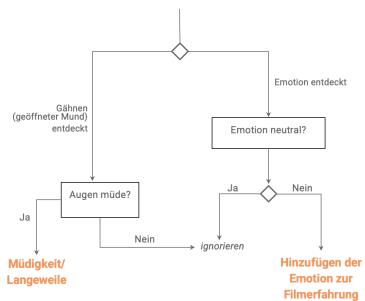


Abbildung 4: Konzept der Integration der Emotions- und Langeweileerkennung in ein System

3.7 Web-Applikation

Die Web-Applikation von *EmotionRate* ist eine zentrale Komponente des Projekts und ermöglicht die Interaktion der Nutzer mit dem System zur Emotionserkennung. Die Applikation ist in *HTML*, *CSS* und *JavaScript* geschrieben und verwendet das *Bootstrap-Framework* für ein responsives Design. Es ist eng mit dem *EmotionRate*-Server verbunden um die Video- und Emotionsdaten dynamisch zu lesen. Um die Erkennung von Emotionen und Langeweile, die mit *Python* implementiert wurden, in eine Web-Applikation zu integrieren, haben wir *Flask* verwendet. *Flask* ist ein Web-Framework für *Python* [1]. Es kann verwendet werden, um einen Webserver auf dem eigenen Computer zu starten. So konnten wir die Python-Komponenten und die Web-Applikation miteinander verbinden. Abbildung 5 veranschaulicht die Architektur unseres Projekts.

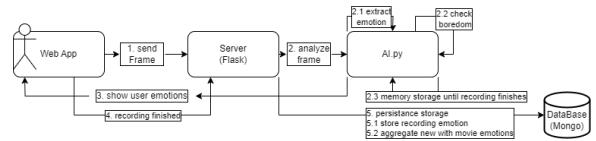


Abbildung 5: Architektur von *EmotionRate*

3.7.1 Implementierung

- *home.html*: Die Startseite bietet eine Übersicht über die verfügbaren Filme und Serien (siehe Abbildung 6). Nutzer können durch die Startseite stöbern, die in einem Layout gestaltet ist, das der Online-Streaming-Plattform Netflix¹ ähnelt. Die Startseite bietet drei Kategorien in Kombination mit einer Infinity-Sroll-Funktion. Ein Navigationsmenü und eine Suchfunktion sind ebenfalls auf der Startseite implementiert.
- *view_clip.html*: Diese Seite zeigt den ausgewählten Film an und visualisiert die über alle Nutzenden aggregierte, allgemeine emotionale Erfahrung des Films oder der Serie (siehe Abbildung 7). Diese wird sowohl durch einen Balken als auch durch detaillierte Prozentangaben veranschaulicht. Der Balken ist dabei in verschiedene Farben unterteilt, wobei jede Farbe den jeweiligen Prozentsatz einer bestimmten Emotion darstellt. Diese allgemeine emotionale Erfahrung wird durch die Aggregation aller erkannten Emotionen und Momente der Langeweile aller Zuschauenden des Films ermittelt, und bietet somit eine umfassende Übersicht. Die Webseite bietet die Möglichkeit, den Film ohne das Starten einer Emotionserkennung anzusehen. Zudem zeigt sie die zugehörigen Filminformationen, zu denen auch die aggregierte

¹<https://www.netflix.com>

emotionale Erfahrung des jeweiligen FilmClips gehört. Ein Frontend-Gerüst für ein Bewertungssystem in Form von Reviews ist ebenfalls implementiert. Die individuelle Emotionserkennung kann durch die Auswahl eines entsprechenden Buttons gestartet werden, wobei der Nutzende zur *new_recording.html* Seite weiterleitet wird.

- *new_recording.html*: Diese Seite ermöglicht es den Nutzenden, eine individuelle Erkennung von Emotionen und Langeweile zu starten (siehe Abbildung 8). Dabei wird das Layout der Seite *view_clip.html* verwendet, das um wichtige Elemente ergänzt wird. Zum einen wird neben dem Filmclip ein Balkendiagramm angezeigt, das die individuell erkannten Emotionen der Zuschauenden visualisiert. Die erkannten Emotionen werden dabei durch den Fortschritt des Balkens visualisiert sowie als Prozentangabe bereitgestellt, die dem Anteil der jeweiligen Emotion an allen erkannten Emotionen entspricht. Unter dem Balkendiagramm befindet sich ein Zähler, der die erkannten Momente der Langeweile anzeigt. Zum anderen wurde ebenfalls ein Video-Feed eingebettet, in dem die Erkennung von Emotionen in Echtzeit beobachtet werden kann. Eine Checkbox erlaubt zudem die Visualisierung der erkannten Mund- und Augenpartie, wie in Kapitel 3.3 erläutert. Das Videomaterial wird mithilfe von *HTML5* und *JavaScript* erfasst und in Echtzeit über die *WebRTC-API* (*Web Real-Time Communication*) an den *EmotionRate*-Server gesendet. Gleichzeitig wird der Filmclip automatisch abgespielt. Die Übertragung erfolgt über eine *WebSocket*-Verbindung, die eine bidirektionale, latenzarme Kommunikation ermöglicht. Die Echtzeit-Videoverarbeitung ist dabei ein zentraler Bestandteil von *EmotionRate*, da sie die Erkennung und Analyse von Emotionen und Langeweile in Echtzeit ermöglicht. Ein Vorteil von *Flask* ist, dass die erkannten Emotionen und Langeweile dynamisch an die Schnittstelle übergeben und die Webseite entsprechend aktualisiert werden kann. Im Frontend werden die Ergebnisse schließlich in Echtzeit visualisiert, sodass Nutzende ihre emotionale Reaktion auf den Film oder die Serie kontinuierlich verfolgen können. Diese Visualisierung erfolgt durch die Aktualisierung des Balkendiagramms zur Anzeige der erkannten Emotionen sowie des Zählers für erkannte Momente der Langeweile. Nach dem Stoppen der Aufnahme werden die individuell erkannten Emotionen und Langeweile ebenfalls in die allgemeine, aggregierte emotionale Erfahrung eines Films integriert. Zudem wird ein Button angeboten, der zur *feedbackForm.html* Seite weiterleitet.

- *feedbackForm.html*: Auf dieser Seite können Nutzende Feedback zur Anwendung bereitstellen. Die Seite enthält ein Feedback-Formular, das für die Studie verwendet wird, um Daten von den Teilnehmenden zu sammeln. Bootstrap-Formular-Komponenten sorgen für eine benutzerfreundliche Eingabe.

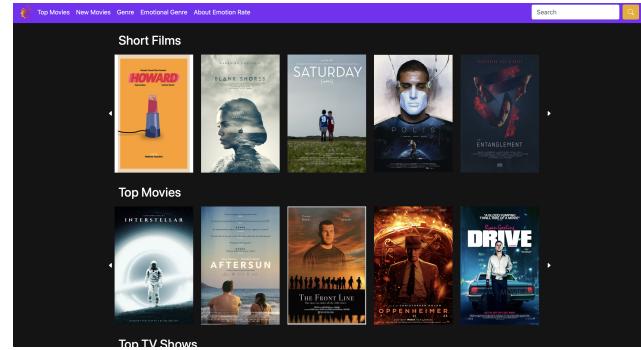
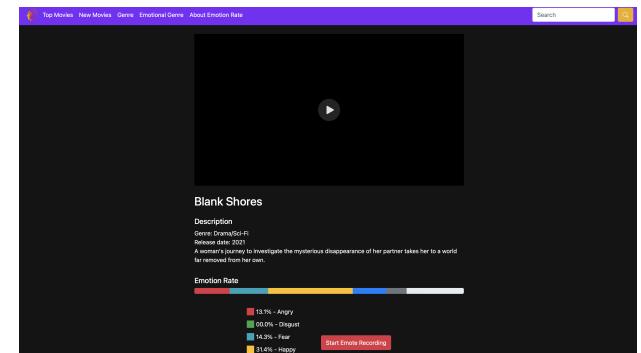
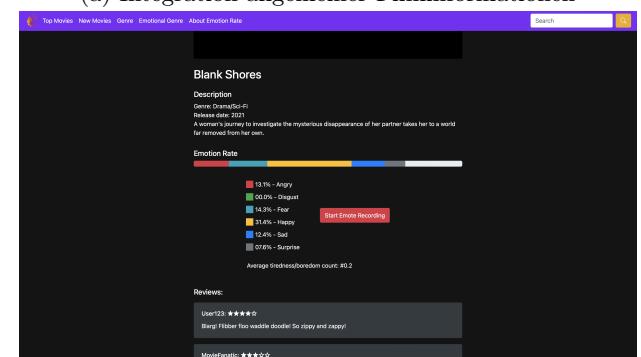


Abbildung 6: Startseite von *EmotionRate*



(a) Integration allgemeiner Filminformationen



(b) Integration der über alle Nutzenden aggregierten, allgemeinen emotionalen Erfahrung eines Films

Abbildung 7: Detailansicht eines Films in *EmotionRate*

Die Hauptfunktionen der Benutzeroberfläche umfassen:

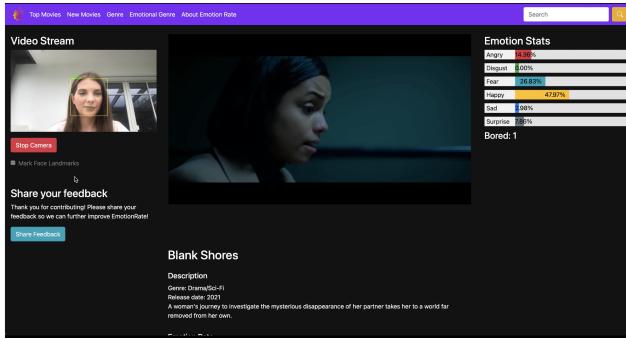


Abbildung 8: Individuelle Erkennung und Visualisierung von Emotionen und Momenten der Langeweile in *EmotionRate*

- Benutzerfreundliche Navigation:** Alle Seiten sind so gestaltet, dass sie eine einfache und intuitive Navigation ermöglichen.
- Echtzeit-Updates:** WebSocket-Verbindungen werden genutzt, um Echtzeit-Updates der Emotions- und Langeweileerkennung anzuzeigen.
- Responsives Design:** Dank Bootstrap passen sich die Seiten an verschiedene Bildschirmgrößen an und bieten ein konsistentes Benutzererlebnis auf Desktops, Tablets und Mobilgeräten.
- Interaktive Elemente:** Formulare, Video-Player und interaktive Diagramme verbessern die Benutzererfahrung und ermöglichen eine direkte Interaktion mit der Anwendung.

3.7.2 Datenbankintegration

EmotionRate integriert eine *MongoDB*-Datenbank, um die erkannten Emotionen und Momente der Langeweile persistent zu speichern. Dies ist notwendig, um die emotionale Erfahrung der Filme und Serien aktuell zu halten. Die Datenbank speichert sowohl die individuell erkannten Emotionen und Momente der Langeweile aller Nutzenden als auch die allgemeine, aggregierte emotionale Erfahrung eines Films oder einer Serie, die aus allen Nutzerdaten berechnet wird. Jede neue, individuelle Emotions- und Langeweilerkennung wird dabei direkt in die allgemeine, aggregierte emotionale Erfahrung eines Films oder einer Serie einbezogen. Darüber hinaus speichert die Datenbank alle integrierten Filme sowie die ausgefüllten Feedback-Formulare der Teilnehmenden, um diese später in der Studie auszuwerten. Über verschiedene API-Endpunkte können Filme sowie emotionale Daten hinzugefügt, abgerufen und aktualisiert werden. Die Datenbankstruktur (Abbildung 9) zeigt die verschiedenen Sammlungen und ihre Beziehungen. Diese Datenbankintegration ermöglicht eine effiziente Speicherung und Abfrage der Daten und unterstützt die nahtlose Integration von Frontend- und Backend-Komponenten.

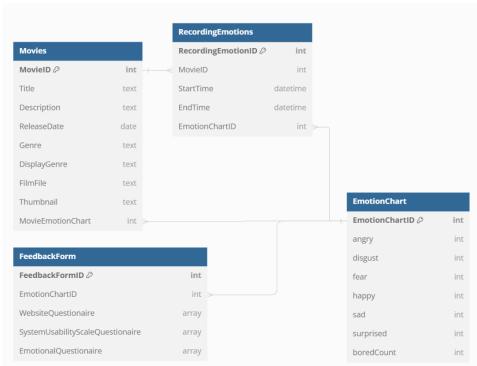


Abbildung 9: Datenbankstruktur

4 Studie

4.1 Studiendesign

Um sicherzustellen, dass *EmotionRate* Emotionen und Langeweile präzise erkennt und um zu evaluieren, ob die Teilnehmenden ein solches Konzept begrüßen würden, wurde eine Studie im Rahmen des Projekts durchgeführt. Der Schwerpunkt lag dabei auf den folgenden beiden Forschungsfragen:

1. Wie kommt *EmotionRate* bei Nutzenden an?
2. Wie häufig stimmt *EmotionRate* mit der Realität überein?

Die erste Forschungsfrage bezieht sich dabei grundlegend auf die Nutzung der Web-Applikation, einschließlich der Sinnhaftigkeit einzelner Elemente und der Benutzerfreundlichkeit der Navigation. Im Gegensatz dazu zielt die zweite Forschungsfrage darauf ab, zu analysieren, ob die tatsächlich empfundene emotionale Erfahrung während eines Films durch die Emotionserkennung angemessen wiedergegeben werden kann. Hierfür wird neben der Analyse der Emotionserkennung zusätzlich eine zweite Methode zur Emotionserkennung verwendet: eine Befragung [21].

4.1.1 Teilnehmende

Die Studie bestand aus 20 Teilnehmenden. Neun von ihnen waren männlich und elf weiblich. Der jüngste Teilnehmende war 17 Jahre alt, der älteste 60 Jahre. Das Durchschnittsalter lag bei 35,3 Jahren. Unter den Teilnehmenden befanden sich 14 Personen mit Hochschulabschluss, fünf mit Fachhochschulreife und eine ohne Abschluss. Darüber hinaus hatten alle 20 Teilnehmenden bereits einmal in ihrem Leben eine Online-Streaming-Plattform von Filmen und Serien genutzt.

4.1.2 Studienablauf

Zunächst wurden die persönlichen Daten der Teilnehmenden erfasst, darunter Geschlecht, Alter, Bildungsabschluss sowie die Angabe, ob sie in ihrem Leben schon einmal eine Online-Streaming-Plattform von Filmen genutzt haben.

Anschließend wurden die Teilnehmenden gebeten, die Web-Applikation explorativ zu erkunden, den Kurzfilm "*Howard*" von Matthew Tompkins aus dem Jahr 2017 anzusehen und schließlich einen Feedback-Fragebogen auszufüllen. Die Erkundung der Web-Applikation erfolgte ohne spezifische Vorgaben, um den Teilnehmenden die Möglichkeit zu geben, sich ein eigenes Bild von der Web-Applikation zu machen. Vor dem Ansehen des Films "*Howard*" wurden die Teilnehmenden darüber informiert, dass sowohl das *Emotion Recording* als auch die Kamera aktiviert werden müssen, um an der Studie teilzunehmen. Dabei wurde ebenfalls betont, den Film im Großbildschirm anzusehen, um sich nicht von der Emotionserkennung ablenken zu lassen. Nach dem Ende des Films, stoppten die Teilnehmenden die Kamera und navigierten problemlos zu dem Feedback-Fragebogen. Dieser wurde anschließend ausgefüllt und abgesendet. Die Feedback-Fragebögen wurden zusammen mit den durch das Modell erkannten Emotionen und Momenten der Langeweile automatisch in der Datenbank gespeichert.

Der Fragebogen umfasste einen allgemeinen Web-Questionnaire, die System-Usability-Scale (*SUS*) nach Brooke [6] sowie eine abschließende Angabe der während des Films empfundenen Emotionen und Langeweile in Prozent. Der Web-Questionnaire beinhaltete Fragen zu Komponenten der Benutzeroberfläche, wie die Navigation, das Farbschemata sowie die Inhaltsrelevanz der Webseite, während der *SUS* eher auf das allgemeine Nutzungserlebnis der Webseite einging. Zusätzlich haben wir neben den sechs Grundemotionen, die unsere Web-Applikation erfasst, auch eine Prozentsangabe für Langeweile aufgenommen. Die Dauer der Studiendurchführung betrug durchschnittlich fünf Minuten und inkludiert die Länge des Films *Howard* von drei Minuten und 13 Sekunden.

4.2 Evaluation

Das folgende Kapitel präsentiert die Ergebnisse bezüglich der quantitativen Daten, der qualitativen Daten sowie den *System Usability Score* der Studie.

4.2.1 Quantitative Daten

Zunächst werden die quantitativen Daten analysiert, um zu evaluieren, welche Emotionen der Film "*Howard*" bei den Teilnehmenden hervorrief und ob er Langeweile verursachte. Dabei wurden die Teilnehmenden angehalten, den Anteil jeder gefühlten Emotion und

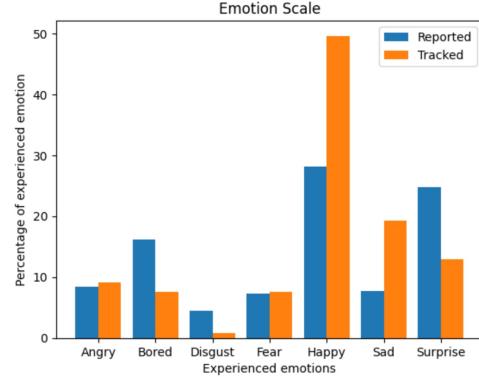


Abbildung 10: Durchschnitt der geloggten und angegebenen Emotionen aller Studienteilnehmenden in Prozent

Langeweile als Prozentsatz aller gefühlten Emotionen anzugeben.

Die Emotion Wut erzielt dabei einen Mittelwert von $\mu = 8,4\%$, die Emotion Abscheu $\mu = 4,4\%$, die Emotion Angst $\mu = 7,3\%$ und die Emotion Traurigkeit $\mu = 7,7\%$. Höherer Prozentsätze erreichten die Emotionen Freude, mit $\mu = 28,2\%$, sowie Überraschung mit $\mu = 24,7\%$. Für Langeweile lag der Mittelwert bei $\mu = 16,1\%$.

Um weiter zu evaluieren, ob die Erkennung der Emotionen und Langeweile mit der Selbstbewertung der Teilnehmenden übereinstimmt, wurden die geloggten Daten zunächst normalisiert. Dabei wurde die Gesamtzahl der Logs für jede erkannte Emotion sowie für Langeweile zur Basis der Normalisierung verwendet. Die normalisierten Daten wurden in Prozentsätzen angegeben, die den Anteil jeder Emotion an der gesamten Anzahl der erfassten Emotionen widerspiegeln. Anschließend wurden die Ergebnisse mit der Selbstbewertung der Teilnehmenden verglichen.

Abbildung 10 zeigt die Gesamtergebnisse aller Emotionen und vergleicht in einem Balkendiagramm den Durchschnitt der durch *EmotionRate* erkannten und der von den Teilnehmenden angegebenen Emotion über alle Studienteilnehmenden in Prozent. Dabei sind die 6 Grundemotionen sowie Langeweile enthalten. Die Bewertung der Emotion *Angst* war dabei in 97,33% der Fälle korrekt. *Wut* wurde zu 92,30% korrekt erkannt, *Freude* zu 56,85%, *Überraschung* zu 52,23% und *Langeweile* zu 47,20%, während die Emotionen *Traurigkeit* zu 39,90% und *Abscheu* zu 18,18% korrekt erkannt wurden.

Darüber hinaus zeigt Abbildung 11 ein Radar-Chart der gemessenen Emotionen von fünf zufällig gewählten Teilnehmenden in Prozent [22]. Es zeigt sich ein deutlicher Trend zu den Emotionen *Freude* und *Traurigkeit*, sowie ein leichter Trend zu der Emotion *Überraschung*.

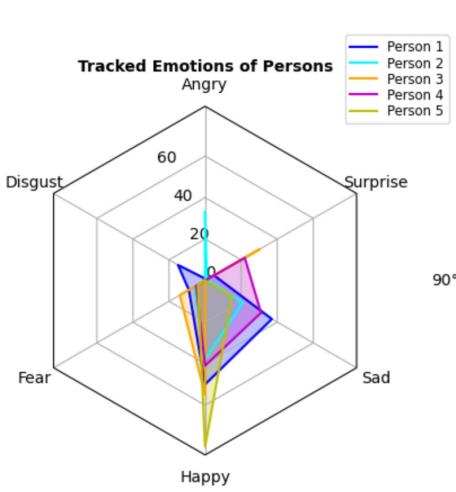


Abbildung 11: Emotionen einzelner, zufällig gewählter Studienteilnehmenden in Prozent

4.2.2 Qualitative Daten

Um unser Konzept zu evaluieren und somit die erste Forschungsfrage zu beantworten, wurden die Teilnehmenden zudem gebeten, einen Web-Questionnaire auf einer fünfstufigen Likert-Skala zu bewerten, die von (1) *Stimme überhaupt nicht zu*, (2) *Stimme nicht zu*, (3) *Neutral*, (4) *Stimme zu*, bis zu (5) *Stimme voll und ganz zu* reichte. Die Ergebnisse sind in Abbildung 12 veranschaulicht.

Zunächst war von Interesse, ob *EmotionRate* eine nutzerfreundliche Schnittstelle integriert. Dabei gaben 100% der Teilnehmenden eine positive Rückmeldung, was in einem Median von $\tilde{x} = 4$ ($min = 4$, $max = 5$) resultiert. Des Weiteren empfanden 95% der Teilnehmenden die Webseite einfach zu navigieren, während 5% der Teilnehmenden neutral eingestellt waren ($\tilde{x} = 4$, $min = 3$, $max = 5$). 90% der Teilnehmenden würden *EmotionRate* weiter empfehlen, wobei 10% der Teilnehmenden eine neutrale Einstellung zeigten ($\tilde{x} = 4$, $min = 3$, $max = 5$). Jeweils 65% der Teilnehmenden empfanden das Farbschema als angenehm ($\tilde{x} = 4$, $min = 2$, $max = 5$), sowie den Inhalt informativ ($\tilde{x} = 4$, $min = 2$, $max = 5$), wobei jeweils 30% der Teilnehmer neutral waren und 5% nicht zustimmten.

4.2.3 System Usability Score

Für eine detaillierte Untersuchung der Usability von *EmotionRate* haben wir die Originalversion der *System Usability Scale* (SUS) verwendet [6]. Diese Skala ist für die Bewertung der Usability diverser Systeme konzipiert und integriert zehn Items, die Einblicke in subjektive Bewertungen der Usability ermöglichen. Für die Bewertung der zehn Items stand den Teilnehmenden eine fünfstufige Likert-Skala von (1) *Stimme überhaupt nicht zu*, (2) *Stimme nicht zu*, (3) *Neutral*, (4) *Stimme zu* bis (5) *Stimme voll und ganz zu* zur Ver-

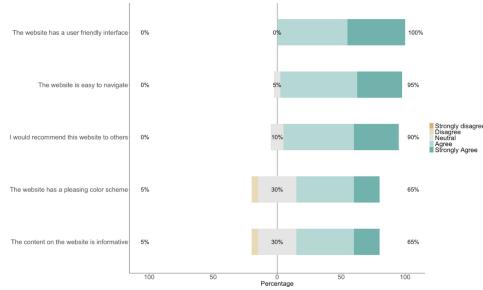


Abbildung 12: Likert-Plot der Ergebnisse des Web-Questionnaires

fügung. Der ermittelte SUS-Score wird anschließend durch eine einzige Zahl im Bereich von 0 (Minimalwert) und 100 (Maximalwert) repräsentiert, die als Maß für die Usability des untersuchten Systems dient. *EmotionRate* erzielte dabei einen Durchschnittswert von $\mu = 93$ (Standardabweichung $\sigma = 11,85$, $min = 72,5$, $max = 100$). Um diese numerische Punktzahl in eine absolute Beurteilung der Usability zu übersetzen, fügten Bangor et al. [4] ein elftes Item als adjektivverankerte Likert-Skala hinzu (*Adjective Rating Scale*), die ebenfalls in Fragebogen integriert wurde. Diese Skala unterstützt die Interpretation der einzelnen SUS-Bewertungen und liefert somit eine subjektive Bezeichnung für den durchschnittlichen SUS-Score. Die Auswertung der Ergebnisse des elften Items implizierte dabei eine *gute* Usability von *EmotionRate*.

5 Diskussion

Die Auswertung der Ergebnisse zeigte, dass die Erfahrungen der Teilnehmenden mit *EmotionRate* im Allgemeinen sehr gut waren, was sich vor allem anhand des insgesamt positiven Feedbacks im Fragebogen feststellen lässt. Die Nutzerfreundlichkeit von *EmotionRate* erhielt eine *gute* Bewertung im SUS. Ebenso kann man beobachten, dass die Teilnehmenden *EmotionRate* empfehlen würden. *EmotionRate* ist intuitiv und einfach nutzbar, was vor allem in der Anlehnung an bereits bekannte Online-Streaming-Plattformen begründet sein kann. Alle Teilnehmenden gaben zu Beginn an, bereits einmal eine solche Plattform genutzt zu haben, weshalb zu vermuten ist, dass sie bereits mit dem Konzept vertraut waren. Anhand der insgesamt deutlich positiven Ergebnisse kann die erste Forschungsfrage bestätigt werden.

Zudem konnte *EmotionRate* die Emotionen der Teilnehmenden gut erkennen und erreichte Genauigkeiten von bis zu 97,33%. Somit kann die zweite Forschungsfrage ebenfalls bestätigt werden. Dennoch unterscheiden sich die Korrektheit bei der Erkennung der Emotionen teilweise von der Selbsteinschätzung der Teilnehmenden. Dafür sind einige Gründe möglich, die einen erheblichen Einfluss auf die Erkennung des Systems ge-

habt haben könnten.

Langeweile kann durch verschiedene Körperteile gezeigt werden [12] und wird nicht immer durch Gähnen oder eine geringe Augenöffnung ausgedrückt, sodass sie seltener von *EmotionRate* erkannt werden kann als sie eigentlich gefühlt wird. In zukünftigen Arbeiten könnte *EmotionRate* somit noch um weitere Indikatoren für Müdigkeit und Langeweile ergänzt werden, wie die Blickrichtung oder spezifische Gesten, wie das Abstützen des Kopfs auf der Hand. Zudem konnten wir feststellen, dass die Sitzposition eine Rolle bei der Erkennung der Augenöffnung spielt. Dies liegt darin begründet, dass Nutzende von unten, oben oder aus einem bestimmten Winkel auf den Bildschirm blicken können, jedoch kann dies die korrekte Erkennung der Augenöffnung beeinträchtigen. Blickt der Nutzende beispielsweise von oben auf den Bildschirm, wird Langeweile schlechter erkannt, da die Augen beim Blick nach oben weiter geöffnet sind. Im Gegensatz dazu, kann Langeweile öfter erkannt werden, wenn der Nutzende leicht von oben auf den Bildschirm blickt, da die Augen hierbei geringer geöffnet sind. Durch eine vorherige Kalibrierung könnte dieses Problem vermieden werden. So könnte das Gesicht zuerst erkannt werden, wenn der Nutzende neutral auf den Bildschirm blickt. Anschließend können Parameter, wie der Schwellenwert der Augenöffnung, individuell angepasst werden.

Darüber hinaus konnte festgestellt werden, dass die Erkennung der Emotion *Abscheu* nur zu 18,18% korrekt war. Diese Unstimmigkeit lässt sich auf die schlechte Erkennung der Emotion aufgrund einer zu geringen Menge an Trainingsdaten im Datensatz zurückführen. Eine Lösung für dieses Problem könnte darin bestehen, die Genauigkeit des CNNs durch weitere Trainingsdaten für die Emotion *Abscheu* zu verbessern.

Zudem zeigte sich, dass die Emotion *Überraschung* öfter gefühlt wurde, als sie von *EmotionRate* erkannt wurde. Der Ausdruck der *Überraschung* ist sehr individuell und daher schwierig zu formalisieren. Oftmals werden ein geöffneter Mund und ein freudiger Blick als *Freude* erkannt. Auf der anderen Seite sind die Emotionen *Freude* und *Traurigkeit* zu häufig analysiert worden. Dabei handelt es sich um Emotionen, die eine Tages-Grundstimmung darstellen können und somit unterbewusst erkennbar sind. Auch hier könnte eine Lösung für dieses Problem darin bestehen, die Genauigkeit des CNN durch ein Finetuning zu optimieren.

Die Verstärkung der Abweichung der Emotion *Freude*, die mit einer Differenz von 21,4 Prozentpunkten von allen Emotionen am größten ist, wird zusätzlich zu Beginn der Emotionserkennung vermehrt ausgelöst. Dies lässt sich vor allem darauf zurückführen, dass Teilnehmende große Freude daran hatten, ihr Gesicht von einer künstlichen Intelligenz analysieren zu lassen. Ebenso können weitere Ablenkungen im Hintergrund als Begründung für die erhöhte Abweichung möglich

sein. Eine größere und längere Studie könnte dieses Problem verringern, da somit Überraschungsmomente einer erstmaligen Nutzung ausgeschlossen werden können. Zudem sollte *EmotionRate* in einer ruhigen Umgebung getestet werden, um Ablenkungen soweit wie möglich zu vermeiden.

Im Allgemeinen lassen sich Abweichungen der erkannten und angebenden Emotionen vor allem dadurch erklären, dass sowohl die Individualität als auch die Ausdruckskraft einer Emotion enorm zwischen unterschiedlichen Charakteren variieren. Diese durch ein standardisiertes Schema zu analysieren ist sehr komplex. Zudem lies sich feststellen, dass beispielsweise die Erkennung von Emotionen durch das Tragen einer Brille stark beeinflusst wurde. So wurden Emotionen durch das Tragen einer Brille seltener erkannt und oftmals falsch interpretiert. Auch hier könnten zusätzliche Trainingsdaten die Genauigkeiten des Modells weiter verbessern.

Hinzugefügt wird eine kurze Betrachtung der Ausreißer. Bei lediglich 20 Teilnehmenden wurden diese in das Ergebnis mit einkalkuliert. Abbildung 13 zeigt eine Einzelperson, deren aufgenommene und angegebene Werte am wenigsten Übereinstimmungen hatten und in einem großen Einfluss auf das Endergebnis resultierten. Die Emotionserkennung hat bei dieser Person kaum mit der Realität übereingestimmt. Abbildung 14 zeigt hingegen eine Einzelperson, deren Werte sehr gut übereinstimmen.

Abbildung 11 lässt sich als genaue emotionale Erfahrung des Films "Howard" interpretieren. Der Trend zu den Emotionen *Freude* und *Traurigkeit*, sowie *Überraschung* lässt sich gut mit dem Genre des Films, einer Komödie, vereinen. Nutzende von *EmotionRate* können somit von unerwarteten emotionalen Situationen während des Films absehen und somit einen passenden Film zu ihre Stimmung auswählen.

Die allgemeinen positiven Ergebnisse bilden eine solide Grundlage, dass es eine große Nachfrage an Online-Streaming-Plattformen mit Emotions- und Langeweile-

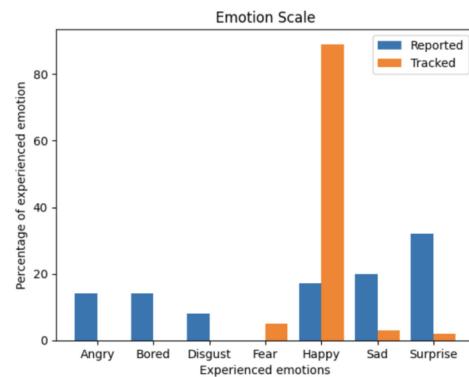


Abbildung 13: Erkannte und angegebene Emotionen und Langeweile einer Person in Prozent

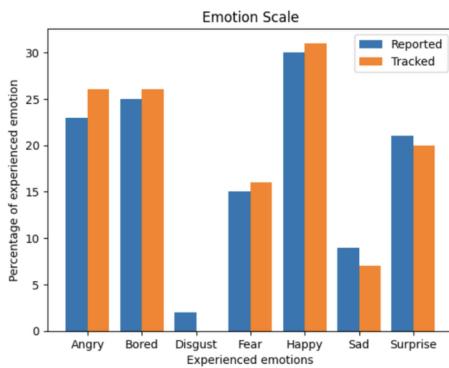


Abbildung 14: Erkannte und angegebene Emotionen und Langeweile einer Person in Prozent

leerkennung gibt. Die Durchführung einer Studie mit einer größeren Anzahl an Teilnehmenden könnte dieser Schlussfolgerung weitere Gültigkeit verleihen.

5.1 Zukünftige Arbeiten

Die Ergebnisse bieten ein großes Potenzial für zukünftige Arbeiten. So könnten zukünftige Parameter miteinbezogen werden, wie beispielsweise die Blickrichtung oder die Kopfneigung, um die Erkennung von Emotionen und Langeweile weiter zu verbessern.

Außerdem lässt sich die erkannte emotionale Erfahrung verbessern, wenn Ablenkungen vermieden werden. Dies würde begünstigen, dass nur Emotionen, die durch den Film ausgelöst werden, in die emotionale Erfahrung miteinbezogen werden. Zudem könnte untersucht werden, ob die Ablenkung ebenfalls als Langeweile gewertet werden kann.

Darüber hinaus könnte erforscht werden, ob die vollständige Erkennung aller Emotionen in der Web-Applikation angezeigt werden sollte. Alternativ könnte diese auf die interessantesten Emotionen, gemessen an Häufigkeiten, beschränkt werden.

Zuletzt war die Studie aufgrund des verfügbaren Zeitrahmens sehr kurz. Eine umfangreiche Studie könnte hierbei weitere Erkenntnisse liefern und die Ergebnisse weiter verfeinern. Auch könnten in einer umfangreicher Studie die Emotionen in Bezug zu der Emotion *Neutral* gesetzt werden, die in diesem Projekt zunächst ignoriert wurde.

6 Fazit

Im Rahmen dieses Projekts wurde ein Prototyp für eine Web-Applikation einer Online-Streaming-Plattform von Filmen und Serien entwickelt, der die Erkennung von Emotionen, sowie Langeweile integriert, um diese als emotionale Erfahrung eines Films oder einer Serie bereitzustellen. Dabei wurde das Ziel verfolgt, die Entscheidungsfindung bei der Film- und Serienauswahl zu erleichtern und die individuelle Stimmung mit der emotionalen Erfahrung eines Films oder einer Serie abstimmen zu können.

Anschließend wurde eine Nutzerstudie durchgeführt. Diese ergab, dass die Genauigkeit der Erkennung von Emotionen und Langeweile von Emotion zu Emotion variiert und Genauigkeiten von 18,18% für die Emotion *Abscheu* bis zu 97,33% für die Emotion *Angst* aufweist. Dennoch ist ein allgemeiner Trend erkennbar, von dem sich eine genaue emotionale Erfahrung ableiten lässt. Die Erkennungsgenauigkeit von *EmotionRate* kann durch eine Anpassung der Parameter durch eine vorherige Kalibrierung und eine Erweiterung des Datensatzes gesteigert werden. Zukünftige Arbeiten können auch darin bestehen, die Anwendung auf Smartphones zu testen.

Im Allgemeinen kam *EmotionRate* sehr gut bei den Teilnehmenden an und erhielt eine gute Bewertung der Usability. Die Mehrheit der Teilnehmenden begrüßte das Konzept von *EmotionRate*. Schließlich war ein hoher Bedarf an einer Online-Streaming-Plattform, die eine Emotions- und Langeweileerkennung integriert und diese als emotionale Filmerfahrung bereitstellt, festzustellen.

Literatur

- [1] What is flask python. Zuletzt besucht: 24.07.2024.
- [2] Takashi Abe. Perclos-based technologies for detecting drowsiness: current evidence and future directions. *Sleep Advances*, 4(1):zpad006, 2023.
- [3] Shaik Afsana. Facial emotion recognizer in real time, 2023.
- [4] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *J. Usability Studies*, 4(3):114–123, may 2009.
- [5] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
- [6] John Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.
- [7] Francois Chollet et al. Keras, 2015.
- [8] Miki Cohen-Kalaf, Joel Lanir, Peter Bak, and Osnat Mokry. Movie emotion map: an interactive tool for exploring movies according to their emotional signature. *Multimedia Tools and Applications*, pages 1–22, 2022.
- [9] John D Eastwood, Alexandra Frischen, Mark J Fenske, and Daniel Smilek. The unengaged mind: Defining boredom in terms of attention. *Perspectives on psychological science*, 7(5):482–495, 2012.
- [10] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.
- [11] Paul Ekman, Robert W Levenson, and Wallace V Friesen. Autonomic nervous system activity distinguishes among emotions. *science*, 221(4616):1208–1210, 1983.
- [12] Meriem El-Yamri, Diego Isar Muñoz, Álvaro Ortiz Marchut, Daniel Padilla Rodríguez, Sofía Prieto Ibañez, José Borja Manero Iglesias, and Josefina Ros Velasco. Vital emo: The boredom detector with a machine learning perspective. 2021.
- [13] Martín Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [14] Nuno Gil, Nuno Silva, Eduardo Duarte, Pedro Martins, Thibault Langlois, and Teresa Chambel. Going through the clouds: search overviews and browsing of movies. In *Proceeding of the 16th International Academic MindTrek Conference*, pages 158–165, 2012.
- [15] James J Gross and Robert W Levenson. Emotion elicitation using films. *Cognition & emotion*, 9(1):87–108, 1995.
- [16] Nafis Irtija, Mahsius Sami, and Md Atiqur Rahman Ahad. Fatigue detection using facial landmarks. In *International Symposium on Affective Science and Engineering ISASE2018*, pages 1–6. Japan Society of Kansei Engineering, 2018.
- [17] Ana Jorge and Teresa Chambel. Visualizations in time for a new look at the movies. *International Journal of Creative Interfaces and Computer Graphics (IJCICG)*, 5(2):40–61, 2014.
- [18] Elnara Kadyrgali, Adilet Yerkin, Yerdauit Torekhan, and Pakizar Shamoi. Group movie selection using multi-channel emotion recognition. *arXiv preprint arXiv:2403.12087*, 2024.
- [19] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.
- [20] Stefan Kroes. Detecting boredom in meetings. *University of Twente: Enschede, The Netherlands*, pages 1–5, 2005.
- [21] Herbert L Meiselman. *Emotion measurement*. Woodhead publishing, 2016.
- [22] Eti Mishra, Piyush Nikam, Sreejith Vidhyadharan, and Reena Cheruvalath. An affect-based approach to detect collective sentiments of film audience: Analyzing emotions and attentions. *Acta Psychologica*, 230:103736, 2022.
- [23] Osnat Mokry, David Bodoff, Nadim Bader, Yael Albo, and Joel Lanir. Sharing emotions: determining films' evoked emotional experience from their online reviews. *Information Retrieval Journal*, 23:475–501, 2020.
- [24] Myriam Munezero, Calkin Suero Montero, Maxim Mozgovoy, and Erkki Sutinen. Emotwitter—a fine-grained visualization system for identifying enduring sentiments in tweets. In *Computational Linguistics and Intelligent Text Processing: 16th International Conference, CICLing 2015, Cairo, Egypt, April 14–20, 2015, Proceedings, Part II 16*, pages 78–91. Springer, 2015.
- [25] Phillip Nelson. Information and consumer behavior. *Journal of political economy*, 78(2):311–329, 1970.
- [26] Michał Ochocki and Dariusz J Sawicki. Yawning recognition based on dynamic analysis and simple measure. In *CHIRa*, pages 111–117, 2017.
- [27] Jonathan Oheix. Face expression recognition dataset, 2018.

- [28] Robert Plutchik. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4):344–350, 2001.
- [29] Yongfeng Qian, Yin Zhang, Xiao Ma, Han Yu, and Limei Peng. Ears: Emotion-aware recommender system based on hybrid information fusion. *Information Fusion*, 46:141–146, 2019.
- [30] James A Russell and Lisa Feldman Barrett. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, 76(5):805, 1999.
- [31] Lea Sigethy, Cristian Gavriliu, Tudor Teofanescu, and Leonie Münster. Gitlab-repository zu emotionrate, 2024.
- [32] Paul J Silvia. Interest and interests: The psychology of constructive capriciousness. *Review of General Psychology*, 5(3):270–290, 2001.
- [33] Ed S Tan. *Emotion and the structure of narrative film: Film as an emotion machine*. Routledge, 2013.
- [34] Jianhua Tao and Tieniu Tan. Affective computing: A review. In *International Conference on Affective computing and intelligent interaction*, pages 981–995. Springer, 2005.
- [35] L Thulasimani, Prithashasni SP, et al. Real time driver drowsiness detection using opencv and facial landmarks. *Int. J. of Aquatic Science*, 12(2):4297–4314, 2021.
- [36] Marko Tkalcic, Andrej Kosir, and Jurij Tasic. Affective recommender systems: the role of emotions in recommender systems. In *Proc. The RecSys 2011 Workshop on Human Decision Making in Recommender Systems*, pages 9–13, 2011.
- [37] Shiliang Zhang, Qi Tian, Qingming Huang, Wen Gao, and Shipeng Li. Utilizing affective analysis for efficient movie browsing. In *2009 16th ieee international conference on image processing (icip)*, pages 1853–1856. IEEE, 2009.
- [38] Zutao Zhang and Jia-shu Zhang. Driver fatigue detection based intelligent vehicle control. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 2, pages 1262–1265. IEEE, 2006.
- [39] Yong Zheng, Bamshad Mobasher, and Robin D Burke. The role of emotions in context-aware recommendation. *Decisions@ RecSys*, 2013:21–28, 2013.
- [40] Qianyang Zhuang, Zhang Kehua, Jiayi Wang, and Qianqian Chen. Driver fatigue detection method based on eye states with pupil and iris segmentation. *Ieee Access*, 8:173440–173449, 2020.