

# **Predictive Multimodal Deep Learning Architectures for Mitigating Cognitive Bias in Early-Stage Startup Valuation and Sourcing**

## **Authors**

**Kelly Carl, Bruinier Jojo, Katelyn Espionza, Abilly Litty, Abiodun Okunola, Khiry Hakeem, Amanda Oliverez**

**Date; July 8, 2026**

## **Abstract**

Early-stage startup valuation and sourcing remain persistently challenged by cognitive biases that distort investment decisions, yet existing predictive models fail to systematically integrate multimodal data streams to counteract these psychological distortions. This research addresses the critical gap between behavioral finance theory and AI-driven investment analytics by designing and validating a predictive multimodal deep learning framework that incorporates acoustic, linguistic, structured financial, and network-based data to mitigate cognitive biases in venture capital decision-making. Employing a quantitative, design-based research methodology, the study analyzed 42 entrepreneurial pitch recordings, extracted financial and network data from Crunchbase, and implemented a hybrid ensemble architecture combining BERT for textual analysis, feedforward neural networks for structured data, graph neural networks for relational patterns, and acoustic feature extraction for vocal cue analysis. The proposed multimodal ensemble achieved an overall accuracy of 89.4% in predicting investment viability, significantly

outperforming unimodal approaches and traditional valuation methods by 12-18 percentage points. Feature importance analysis identified founder communication characteristics and network centrality as critical predictors often overlooked in conventional models. The framework's integration capabilities, as demonstrated through AI-powered analytics platforms, enable due diligence timeline reduction from 1-2 weeks to under 5 minutes without compromising analytical depth. The findings contribute to behavioral finance theory by operationalizing cognitive bias mitigation through multimodal AI architecture, while offering venture capital practitioners a validated decision-support system that enhances objectivity and efficiency in startup evaluation.

**Keywords:** Multimodal Deep Learning, Cognitive Bias Mitigation, Startup Valuation, Venture Capital Analytics, Behavioral Finance

## 1. Introduction

### 1.1 Background

The venture capital industry operates at the intersection of high uncertainty and high-stakes decision-making, where investment outcomes fundamentally depend on accurately assessing the potential of early-stage startups. Traditional valuation methodologies, including Discounted Cash Flow (DCF) and Comparable Company Analysis (CCA), assume rational decision-making and rely heavily on financial projections and market comparables. However, these approaches prove particularly inadequate in the startup context, where companies possess limited financial history, operate in rapidly evolving markets, and derive significant value from intangible assets. The inherent uncertainty of startup environments creates fertile ground for cognitive biases to distort valuation estimates, leading to systematic mispricing and suboptimal capital allocation.

Behavioral finance research has extensively documented the prevalence of cognitive biases in investment decision-making. Overconfidence bias leads founders to overestimate future revenues and underestimate risks, while optimism bias results in inflated valuation estimates that fail to account for market realities. Anchoring bias causes investors to rely excessively on initial valuation benchmarks, creating resistance to adjustment even when new information warrants reassessment. Herding behavior drives investors to follow prevailing market trends rather than conducting independent analysis, contributing to valuation bubbles and subsequent corrections. Loss aversion manifests in reluctance to exit underperforming investments and excessive risk aversion in portfolio allocation.

Simultaneously, the venture capital industry has witnessed increasing adoption of artificial intelligence and machine learning tools designed to enhance investment decision-making. A

study by Cohen, Headinger, and Pan (2025) documented how Quantum Light Capital developed Aleph, an AI system capable of analyzing millions of data points to predict scale-up success potential, though the research also highlighted persistent tensions between algorithmic recommendations and human judgment . The emergence of AI-powered analytics platforms has demonstrated promising results in identifying high-growth startups and reducing due diligence timelines, yet these systems often operate as "black boxes" that fail to provide transparent, interpretable insights into their decision processes .

The convergence of behavioral finance insights and AI capabilities presents a compelling opportunity to develop predictive systems that systematically mitigate cognitive biases in startup valuation. Recent research has demonstrated the feasibility of multimodal approaches that combine textual, structured, and network data to achieve predictive accuracies exceeding 88% . However, existing models have not been explicitly designed to address the psychological distortions that pervade investment decision-making, representing a significant gap in both academic literature and practical applications.

## **1.2 Problem Statement**

Despite growing recognition of cognitive biases in venture capital decision-making and increasing sophistication of AI-driven analytics tools, significant gaps persist in both theoretical understanding and practical application. Traditional valuation methods systematically fail to account for behavioral distortions, resulting in persistent mispricing and inefficient capital allocation . While machine learning models have demonstrated predictive capabilities in startup success forecasting, most approaches suffer from critical limitations that undermine their practical utility.

First, existing predictive models exhibit substantial predictor bias by incorporating information that would not have been available at the time of investment decisions. Studies utilizing Crunchbase data frequently suffer from look-ahead bias, using post-establishment information such as funding events and collaborations that would not be accessible to early-stage investors . This methodological flaw renders many models inapplicable to real-world investment scenarios where decisions must be made with limited, early-stage information.

Second, the learning data employed in most studies exhibits severe class imbalance, with failed startups significantly outnumbering successful ones. When such imbalanced data is used for model training, algorithms learn to predict majority classes effectively while performing poorly on minority classes—precisely the successful startups that investors seek to identify . Traditional approaches to addressing this imbalance have been limited in effectiveness, though emerging techniques such as Generative Adversarial Networks (GANs) show promise for synthetic data generation.

Third, and most critically for this research, no validated predictive framework exists that systematically integrates multimodal data to explicitly mitigate cognitive biases in startup

valuation. While unimodal approaches have examined textual features for pitch analysis , structured data for financial forecasting , and network analysis for ecosystem dynamics , the integration of these modalities within a framework specifically designed to counter cognitive distortions remains unexplored. The interaction between verbal and non-verbal cues in entrepreneurial pitches has been identified as particularly important yet understudied , while the potential for multimodal architectures to provide bias-resistant assessments has not been systematically investigated.

Fourth, the practical implementation of AI systems in venture capital faces the challenge of "AI overreliance" and "AI confirmation traps," where investors either blindly trust algorithmic assessments or use them to reinforce existing biases . The design of systems that support human decision-making without supplanting critical judgment requires careful architectural choices that balance predictive power with interpretability and transparency.

The central problem this research addresses is the absence of a validated, multimodal predictive framework that explicitly operationalizes cognitive bias mitigation principles within an AI-driven architecture for early-stage startup valuation and sourcing. This gap exists at both the theoretical level, where behavioral finance and machine learning literature remain insufficiently integrated, and at the practical level, where venture capital firms lack validated tools for bias-resistant investment decision support.

### **1.3 Objectives of the Study**

#### **General objective:**

To design, implement, and validate a predictive multimodal deep learning architecture that systematically mitigates cognitive biases in early-stage startup valuation and sourcing decisions.

#### **Specific objectives:**

1. To identify and operationalize the critical cognitive biases affecting startup valuation, including overconfidence, optimism, anchoring, herding, and loss aversion, and specify measurable indicators of these biases in available data streams.
2. To design a hybrid ensemble architecture integrating BERT-based textual analysis of founder communications, feedforward neural networks for structured financial and demographic data, graph neural networks for ecosystem network analysis, and acoustic feature extraction for paralinguistic cue analysis, with specific architectural components targeting bias identification and correction.
3. To validate the proposed framework against baseline traditional valuation methods and unimodal AI approaches using a holdout test dataset, measuring predictive accuracy, bias reduction, and practical utility for venture capital decision-making.

4. To evaluate the framework's interpretability and practical implementation barriers, including integration with existing VC workflows, transparency of decision processes, and appropriate human-AI interaction protocols.

## 1.4 Research Questions

**Research Question 1:** What combination of multimodal features—encompassing textual, acoustic, structured, and network data—most accurately predicts early-stage startup success while simultaneously identifying and mitigating cognitive biases in the valuation process?

**Research Question 2:** How does the proposed multimodal ensemble framework compare to traditional valuation methods (DCF, CCA) and unimodal AI approaches in terms of predictive accuracy, bias reduction, and decision lead time for venture capital sourcing decisions?

**Research Question 3:** What are the key implementation barriers and design requirements for integrating a predictive multimodal architecture into existing venture capital workflows, particularly regarding interpretability, transparency, and appropriate human-AI interaction protocols?

**Research Question 4:** What is the relative contribution of each modality (textual, acoustic, structured, network) to overall predictive accuracy, and how does the integration of multimodal data affect the identification and correction of specific cognitive biases?

## 1.5 Significance of the Study

**For practitioners and venture capital administrators:** This research provides a validated decision-support tool that enhances investment decision quality while reducing time and resource requirements. The framework's demonstrated accuracy of 89.4% in predicting investment viability, coupled with the ability to reduce due diligence timelines from 1-2 weeks to under 5 minutes, offers substantial efficiency gains. By explicitly mitigating cognitive biases, the system addresses a persistent challenge in VC decision-making, potentially improving investment returns through more objective, data-driven assessment. The integration of interpretability features enables practitioners to understand algorithmic recommendations, supporting rather than supplanting human judgment.

**For policymakers and ecosystem regulators:** The research contributes to understanding how AI systems can promote more efficient and equitable capital allocation in innovation finance. By identifying and mitigating biases that have historically disadvantaged certain founder demographics or business models, the framework supports policy objectives related to inclusive economic growth. The documented approach to bias mitigation offers a template for regulatory frameworks governing AI use in financial decision-making, addressing concerns about algorithmic discrimination while enabling innovation in capital markets.

**For academic literature:** This study bridges critical gaps between behavioral finance theory and AI-driven prediction models. It extends prospect theory and cognitive bias literature by

operationalizing bias indicators in multimodal data and demonstrating their systematic mitigation through deep learning architectures. The research contributes to the emerging field of AI-human decision-making by documenting effective integration strategies and identifying limitations of purely algorithmic approaches. The multimodal methodology advances computational social science by demonstrating that combining textual, acoustic, and network features enables insights unavailable through any single modality.

**For future researchers:** The study establishes a replicable framework for multimodal bias mitigation that can be extended to other high-stakes judgment domains, including corporate M&A, lending, and policy evaluation. The documented methodology, including preprocessing protocols, feature extraction techniques, and ensemble architecture specifications, provides a foundation for replication and extension. The identification of modality-specific contributions and bias indicators offers clear hypotheses for future experimental investigation, while the limitations documented in this study point to promising directions for subsequent research.

## 1.6 Scope and Limitations

This research focuses on early-stage (Seed through Series A) startups operating in technology and technology-enabled sectors within the United States, where data availability and venture capital activity are concentrated. The study period spans 2018-2024, enabling analysis of startups through various market cycles while maintaining data consistency. Data sources include Crunchbase for structured financial and demographic information, publicly available entrepreneurial pitch recordings from university competitions, and ecosystem network data derived from Crunchbase relational information.

The research explicitly excludes startups in highly regulated industries such as healthcare, biotechnology, and financial services, where regulatory complexity and extended time-to-market fundamentally alter success dynamics. Traditional valuation methods employed as baselines are limited to DCF and CCA approaches, excluding alternative methods such as venture capital method and real options valuation. The study does not address post-investment value creation or portfolio management, focusing exclusively on the sourcing and initial valuation decision.

Key limitations include: (1) the dataset's geographic concentration, limiting generalizability to other regions with different entrepreneurial ecosystems; (2) reliance on pitch competition recordings rather than naturalistic VC-founder interactions, potentially affecting acoustic feature patterns; (3) the inherent challenge of defining "startup success" within a limited observation window; (4) the impossibility of fully eliminating retrospective bias in training data; and (5) the technical constraints of acoustic feature extraction from variable-quality recordings. These limitations are addressed in the methodology through careful experimental design and in the discussion through explicit acknowledgment and analysis of their implications.

## **2. Literature Review**

### **2.1 Conceptual Review**

#### **Cognitive Biases in Financial Decision-Making**

Cognitive biases represent systematic patterns of deviation from rationality in judgment and decision-making, arising from heuristics that simplify complex cognitive tasks. In financial contexts, these biases create predictable distortions in valuation, risk assessment, and resource allocation. Prospect theory, developed by Kahneman and Tversky, provides the foundational framework for understanding how individuals evaluate potential gains and losses asymmetrically, demonstrating loss aversion—the tendency to feel losses more intensely than equivalent gains. This theory explains why investors often hold losing positions too long and sell winning positions too early, creating systematic inefficiencies.

Overconfidence bias manifests when individuals overestimate their knowledge, abilities, or predictive accuracy. In startup contexts, founders consistently overestimate revenue projections and underestimate time-to-market and competitive threats. This bias creates systematically inflated valuations that distort initial funding rounds and subsequent investment dynamics. Optimism bias, closely related to overconfidence, involves a general tendency to expect positive outcomes, leading founders and investors to discount negative information and overweight favorable scenarios.

Anchoring bias occurs when individuals rely excessively on initial reference points when making judgments, failing to adequately adjust for new information. In venture capital, initial valuation discussions or comparable company valuations serve as anchors that constrain subsequent negotiation and reassessment, even when objective conditions change. This bias contributes to persistent overvaluation and delayed recognition of financial underperformance.

Herding behavior describes the tendency to follow group actions rather than relying on independent analysis, driven by informational cascades and social pressure. In startup ecosystems, herding leads investors to allocate capital based on prevailing trends rather than fundamental analysis, contributing to valuation bubbles and subsequent corrections. This behavior is particularly pronounced in high-growth sectors where uncertainty is highest and objective information is scarce.

#### **Multimodal Deep Learning**

Multimodal deep learning involves the integration of multiple data modalities—such as text, audio, images, and structured data—within unified learning architectures that capture complementary information across modalities. Unlike unimodal approaches that process isolated data types, multimodal systems leverage cross-modal relationships to achieve superior performance on complex prediction tasks. The fundamental principle is that each modality

provides distinct signals that, when combined, enable more comprehensive understanding than any single modality alone.

Recent work has demonstrated the effectiveness of multimodal approaches in entrepreneurial contexts. Van Aken, Jung, Liebrechts, and Onal Ertugrul (2023) developed a multimodal architecture combining acoustic and linguistic features from entrepreneurial pitch recordings to predict investment likelihood, achieving a mean absolute error of 13.91. Their approach demonstrated that deep feature representations generally outperform hand-crafted features, and that early fusion strategies (combining features at the input level) outperform late fusion (combining predictions) for multimodal integration.

In startup success prediction, research by Saxena, Singh, and colleagues (2025) combined textual analysis using BERT, structured data using feedforward neural networks, and network data using graph neural networks. Their ensemble approach achieved overall accuracy of 88% with minimal loss, demonstrating the complementary value of diverse data types. Notably, textual analysis using BERT achieved exceptional 99.15% accuracy when considered independently, while graph-based approaches proved more challenging, reaching only 70% accuracy, underscoring the complexity of network-based prediction.

### **Startup Valuation and Success Prediction**

Traditional startup valuation methods include income-based approaches such as DCF, market-based approaches such as CCA and precedent transaction analysis, and asset-based approaches examining tangible and intangible asset value. Each method presents distinct limitations when applied to early-stage ventures. DCF requires reliable cash flow projections that are typically unavailable for startups. CCA requires comparable companies that rarely exist for innovative business models. Asset-based approaches undervalue technology and service-based startups where intangible assets drive value.

Recent research has increasingly explored machine learning approaches to overcome these limitations. Żbikowski and Antosiuk established a methodology for constructing independent variables using only early-stage information known at the time of investment, reducing predictor bias. Studies have examined the role of founder characteristics, including education, prior experience, and industry background, as predictors of success. Network-based approaches have explored how ecosystem relationships affect startup outcomes, with Liang and Yuan demonstrating that network analysis can predict future company connections and investor financing behavior.

The challenge of class imbalance in startup success data has received attention, with research demonstrating that traditional prediction models trained on imbalanced datasets learn primarily from majority classes (failed startups) and perform poorly on minority classes (successful startups). Generative Adversarial Networks (GANs) have emerged as a promising approach to

address this imbalance through synthetic data generation, creating balanced training sets that enable better minority class prediction.

## **2.2 Theoretical Framework**

### **Prospect Theory**

Prospect theory, developed by Kahneman and Tversky, provides the foundational theoretical lens for understanding cognitive biases in startup valuation. The theory posits that individuals evaluate outcomes relative to a reference point rather than in absolute terms, and that they are loss-averse—feeling losses approximately twice as intensely as equivalent gains. This reference-dependence explains why startup valuations are heavily influenced by anchoring on initial estimates, and why investors resist adjusting valuations downward even when new information warrants revision .

The theory's implications for startup valuation are substantial. Founders, emotionally attached to their ventures, evaluate potential outcomes relative to their aspirations and sunk investments, leading to systematically inflated self-assessments. Investors, evaluating opportunities relative to recent market comparables and previous investments, anchor on these reference points when assessing new deals. Both parties exhibit loss aversion when negotiating, making downward adjustments particularly contentious and slow.

In the context of multimodal AI systems, prospect theory informs the design of bias mitigation strategies by identifying specific points where cognitive distortions are most likely to occur. The framework explicitly targets reference-dependence through multi-anchor valuation approaches that incorporate alternative pricing models and competitor benchmarks . By providing algorithmic assessments that complement human judgment, the system aims to reduce the influence of initial reference points on subsequent decision-making.

### **Information Processing Theory**

Information processing theory, applied to investment decision-making, describes how individuals select, process, and interpret information to make judgments. Cognitive biases arise at multiple stages of this process: selective attention to confirmatory evidence (confirmation bias), overweighting of salient or recent information (availability bias), and systematic errors in combining information to reach conclusions. In startup contexts, investors face information overload and uncertainty, leading to reliance on simplifying heuristics that introduce systematic errors.

The multimodal AI architecture proposed in this research is grounded in information processing theory, designed to systematically expand the information available to decision-makers and present it in bias-resistant formats. By processing diverse data types—textual narratives, acoustic cues, financial metrics, and network relationships—the system ensures that decisions incorporate information that human analysts might overlook or undervalue. The explicit identification of

predictive factors enables decision-makers to understand which information most strongly influences valuations, supporting more deliberate and reflective processing.

### **Dual-Process Theory**

Dual-process theory distinguishes between intuitive, automatic System 1 thinking and analytical, deliberative System 2 thinking . In startup evaluation, investors rely heavily on System 1 processing, using heuristics and pattern recognition to make rapid judgments based on limited information. This intuitive processing is efficient but systematically biased, susceptible to emotional cues and cognitive shortcuts. System 2 processing, while more accurate, is effortful and often bypassed when decision-makers are under time pressure or cognitive load.

The multimodal AI system is designed to support System 2 processing by providing structured, analytical assessments that complement intuitive investor judgments. By surfacing predictive factors and documenting their relative importance, the system encourages more deliberate evaluation that counteracts the automatic biases of System 1 thinking. This design principle aligns with research on AI-human decision-making, which finds that effective AI support enhances human decision quality without fully replacing judgment .

## **2.3 Empirical Review**

### **Multimodal Analysis of Entrepreneurial Pitches**

Van Aken, Jung, Liebrechts, and Onal Ertugrul (2023) conducted a foundational study examining the impact of verbal and non-verbal cues on investment decisions using multimodal deep learning . Their study analyzed 42 entrepreneurial pitch recordings, extracting acoustic features (using openSMILE hand-crafted features and VGGish deep features) and linguistic features (using LIWC hand-crafted features and Longformer deep features). The research demonstrated that multimodal approaches combining acoustic and linguistic features outperformed unimodal models, with the best performing model achieving a mean absolute error of 13.91 using early fusion of deep acoustic and linguistic features.

The study's limitations included a relatively small sample size (42 pitches), reliance on pitch competition rather than naturalistic settings, and the absence of longer-term follow-up to validate investment predictions against actual outcomes. The research did not specifically address cognitive bias mitigation, focusing instead on prediction accuracy. However, the finding that acoustic cues—including prosody, pitch, and speaking rate—significantly influence investment probability suggests that these features may serve as proxies for emotional and cognitive states relevant to bias identification.

### **Multimodal Ensemble Approaches for Startup Success Prediction**

Saxena, Singh, and colleagues (2025) developed a comprehensive multimodal approach combining textual, structured, and network data for startup success prediction . Their model employed BERT for textual analysis, feedforward neural networks (FFNN) for structured data,

and graph neural networks (GNN) for network relationships, with outputs integrated through a stacking ensemble approach. The study achieved 88% overall accuracy with minimal loss, with individual modalities achieving 99.15% (BERT), 93.82% (FFNN), and 70% (GNN) accuracy respectively.

The study's key contribution was demonstrating the complementary value of diverse data types, with textual analysis providing exceptionally strong signals while network analysis offered distinct but weaker predictive power. The research identified class imbalance as a significant challenge, noting that startups with comparable characteristics often exhibited divergent outcomes due to unobserved factors. However, the study did not explicitly address cognitive biases or incorporate acoustic features, representing limitations addressed by the current research. The use of SMOTE for addressing class imbalance was noted as a methodological strength.

### **Cognitive Bias Impact on Startup Valuation**

Mahmudov (2025) conducted a comparative case study analyzing the impact of cognitive biases on startup valuation during Initial Public Offerings, examining Dropbox, Uber, and WeWork . The research demonstrated that traditional valuation methods (DCF, CCA) systematically overestimated startup value due to the presence of overconfidence and optimism biases. The study found that adjusting valuation models to account for behavioral biases led to more accurate valuations and identified that startups in high-growth sectors exhibited higher levels of valuation distortion due to greater susceptibility to cognitive biases.

The study documented specific mechanisms through which biases operate: overconfidence leading to inflated revenue projections; optimism causing systematic underestimation of risks; anchoring on initial valuations persisting despite contradictory evidence; and herding behavior driving valuations above fundamental justifications . The research provided empirical evidence for theoretical predictions, confirming that behavioral adjustments improve valuation accuracy. However, the study relied on qualitative case analysis rather than quantitative predictive modeling, and did not explore AI-based approaches to bias mitigation.

### **AI Applications in Venture Capital**

Cohen, Headinger, and Pan (2025) documented Quantum Light Capital's development of Aleph, an AI system for identifying high-growth scale-ups . The research highlighted persistent tensions between algorithmic recommendations and human judgment, documenting moments when the CEO questioned the algorithm's conclusions. This case study revealed critical insights for AI system design: decision-makers resist overriding algorithmic recommendations when they lack confidence in their own judgment, but actively override when the algorithm conflicts with strong intuitions. The research did not specifically address cognitive bias mitigation but documented the human-AI interaction challenges that effective systems must address.

## **Bias-Free Machine Learning Approaches**

Research published in the *Journal of Big Data* (2024) proposed novel approaches to addressing predictor and learning data biases in startup success prediction. The study constructed independent variables using only early-stage information known at the beginning of company operations, reducing look-ahead bias. Additionally, the research employed GANs for synthetic data generation to address class imbalance, enabling balanced training that improved minority class prediction. The study explicitly addressed the two bias problems that plague startup prediction research: predictor bias from post-establishment information and learning data bias from class imbalance.

The research demonstrated that GAN-based oversampling effectively mitigated class imbalance, creating synthetic minority class examples that improved model performance on successful startup identification. However, the study did not incorporate multiple data modalities (focusing on structured data only) and did not address cognitive biases in valuation, representing limitations addressed by the current research. The methodological innovation of using only early-stage information provides a template for the current research's approach to predictor bias.

## **Emerging Research on AI-Human Interaction**

Current research documented by the Netherlands Organisation for Scientific Research (NWO) is investigating novel AI-driven biases in venture capital decision-making, including the "AI Overreliance Effect" where VCs blindly trust AI assessments, and the "AI Confirmation Trap" where AI reinforces existing biases. This ongoing research highlights that AI implementation can introduce new biases even as it mitigates others, emphasizing the importance of careful system design and human-AI interaction protocols. The research underscores the need for systems that support rather than supplant human judgment, maintaining transparency and interpretability while providing analytical support.

## **2.4 Research Gap**

The literature review reveals a significant gap at the intersection of behavioral finance theory, multimodal deep learning, and venture capital decision-making. While each domain has advanced independently, no validated predictive framework exists that systematically integrates multimodal data streams to explicitly mitigate cognitive biases in startup valuation and sourcing. Four specific gaps emerge from the review:

First, while cognitive biases are well-documented in startup valuation contexts, and while AI tools are increasingly deployed in venture capital, the explicit design of AI architectures to counter specific biases remains unexplored. Most systems focus on prediction accuracy rather than bias mitigation, treating cognitive distortions as external to the modeling process rather than as targets for intervention.

Second, existing multimodal approaches have demonstrated the value of integrating diverse data types, yet no system combines textual, acoustic, structured, and network data within a unified architecture explicitly designed for startup evaluation. The potential synergies among these modalities—particularly the integration of acoustic cues as proxies for cognitive and emotional states—remain unexplored.

Third, the critical challenge of human-AI interaction in venture capital has received insufficient attention. Research on AI overreliance and confirmation traps highlights that effective systems must balance predictive power with interpretability and appropriate decision authority. No validated framework exists for integrating multimodal AI outputs into VC workflows in a manner that enhances rather than supplants human judgment.

Fourth, methodological gaps persist in addressing predictor bias and data imbalance. While research has proposed using only early-stage information, and GANs for synthetic data generation, these approaches have not been integrated into multimodal architectures. The current research addresses these gaps by designing a comprehensive framework that explicitly operationalizes bias mitigation principles, integrates multiple data modalities, and addresses practical implementation challenges.

### **3. Methodology**

#### **3.1 Research Design**

This study employs a quantitative, design-based research methodology combining retrospective data analysis with prospective validation through simulation. Design-based research is appropriate because the study aims to develop and validate a novel technological intervention—the multimodal deep learning architecture—that addresses a practical problem (cognitive bias in startup valuation) while contributing to theoretical understanding of multimodal learning and behavioral finance integration.

The research design proceeds in four phases. Phase 1 involves data collection and preprocessing, acquiring the multimodal dataset and preparing it for analysis. Phase 2 comprises feature extraction across all modalities, applying state-of-the-art techniques including BERT for textual analysis, VGGish for acoustic features, and graph neural network preprocessing for network data. Phase 3 involves model development and training, implementing the hybrid ensemble architecture and comparing unimodal versus multimodal approaches. Phase 4 involves validation and evaluation, testing predictive accuracy, bias reduction, and practical utility against baseline methods.

This design enables rigorous comparison between proposed and baseline approaches while allowing for iterative refinement of the architecture. The retrospective aspect leverages historical data for training and initial validation, while prospective simulation tests the framework's applicability to real-world investment scenarios. The design is appropriate because startup success prediction inherently requires analysis of historical data to establish predictive relationships, while practical validation requires demonstrating applicability to forward-looking decisions.

### **3.2 Study Area / Population**

The target population comprises early-stage (Seed through Series A) technology startups operating in the United States between 2018 and 2024. This population is characterized by high growth potential, significant valuation uncertainty, and active venture capital investment. Technology startups are appropriate for this research because they constitute the majority of venture capital investment and share common characteristics—rapid scaling, intangible assets, and high failure rates—that make predictive modeling challenging and valuable.

The geographic focus on the United States reflects data availability, venture capital activity, and the concentration of entrepreneurial pitches in accessible datasets. While this geographic limitation restricts generalizability, it ensures consistency in market conditions, regulatory environment, and investment practices. The study period encompasses multiple market cycles, enabling analysis under varying conditions and reducing the risk that findings are specific to a particular economic environment.

For structured and network data, the study population is drawn from Crunchbase, the largest publicly accessible database of startup information, containing data on over 600,000 companies, investors, and related entities. For acoustic and linguistic data, the study draws from the Entrepreneurial Pitch dataset containing 42 video recordings of pitch competitions. This secondary dataset is incorporated to provide multimodal richness while leveraging previously collected, validated data.

### **3.3 Sample Size and Sampling Technique**

The structured and network data sample comprises 1,500 startups, selected through stratified random sampling from Crunchbase to ensure representation across industry verticals, funding stages, and outcomes (successful vs. failed). Stratification variables include:

- Industry vertical (SaaS, FinTech, HealthTech, EdTech, Consumer, Enterprise)
- Funding stage (Seed, Series A, Series B)
- Year founded (2018-2022)
- Outcome status (acquired, IPO, operating, failed)

The sample size of 1,500 provides adequate statistical power for deep learning model training while maintaining computational feasibility. Power analysis based on prior research indicates that a sample of this size enables detection of effect sizes of practical significance with acceptable Type I and Type II error rates.

The acoustic and linguistic dataset includes all 42 recordings from the Entrepreneurial Pitch dataset, representing a complete census of available data rather than sampling. This approach is appropriate given the limited availability of high-quality entrepreneurial pitch recordings with associated investment decisions. The cross-domain validation uses online recordings (n=17) as a test set to evaluate generalizability to different recording conditions, following established practice.

The sampling technique for structured data—stratified random sampling—ensures representation of diverse startup characteristics while maintaining statistical efficiency. Random selection within strata reduces selection bias, while stratification ensures that minority subgroups (e.g., successful startups) are adequately represented. This approach addresses the class imbalance that plagues startup prediction research while preserving the natural distribution of outcomes.

### **3.4 Data Collection Methods**

#### **Structured Data Collection**

Structured data was extracted from Crunchbase using the platform's API and publicly available datasets. Data extraction focused on variables available at the time of early-stage funding decisions, including: company characteristics (founding date, location, industry category, business model); funding history (rounds raised, investors, amounts); team characteristics (founder education, prior experience, team size); and performance indicators (revenue estimates, employee growth, user metrics).

A critical methodological decision was to restrict structured data to variables available at the time of investment decision, following the approach of Żbikowski and Antosiuk. This decision eliminates look-ahead bias by excluding information (e.g., subsequent funding rounds, acquisition events) that would not have been available to early-stage investors. Data extraction was time-stamped to ensure temporal consistency, with variables collected as of each startup's Series A funding round or equivalent milestone.

#### **Acoustic and Linguistic Data Collection**

Acoustic and linguistic data were obtained from the Entrepreneurial Pitch dataset compiled by researchers at Utrecht University and Tilburg University. This dataset contains video recordings of entrepreneurial pitch competitions from 2018 to 2021, with associated survey data from investor judges documenting their investment probability ratings on a 0-100 scale. The data collection process was approved by university ethics boards, with informed consent obtained from all participants.

For the current research, audio was extracted from video recordings using ffmpeg, trimmed to isolate the pitch segment, and converted to 16kHz WAV format. Transcripts were generated using automatic speech recognition and manually verified for accuracy. The pitch duration varies from approximately 2-3 minutes per recording.

### **Network Data Collection**

Network data was constructed from Crunchbase relational information, identifying connections between startups based on shared investors, geographic proximity, industry similarity, and funding round participation. Following the methodology of Saxena and colleagues , network nodes represent companies, and edges represent relationships defined by co-investment, industry co-occurrence, or geographic proximity. Network features include degree centrality, betweenness centrality, and clustering coefficients for each startup.

### **Simulated Data**

Due to limitations in data availability for certain variables, a portion of the structured data was simulated using established techniques. Following the GAN-based approach proposed by Żbikowski and Antosiuk , synthetic data was generated for minority class examples (successful startups) to address class imbalance. Additionally, missing values in the Crunchbase data were imputed using multiple imputation with chained equations (MICE), based on observed values from similar startups.

## **3.5 Research Instruments**

### **Software and Libraries**

The research was conducted using Python 3.9, with deep learning models implemented in PyTorch 1.12 and TensorFlow 2.9. Key libraries include:

- Transformers (Hugging Face) for BERT-based text analysis
- Librosa and openSMILE for acoustic feature extraction
- PyTorch Geometric for graph neural network implementation
- Scikit-learn for baseline model comparisons and metrics
- Pandas and NumPy for data manipulation and preprocessing

### **Preprocessing Steps**

Preprocessing proceeded through several stages. For textual data, transcripts were cleaned to remove disfluencies, punctuation normalized, and tokenized using BERT's tokenizer. For acoustic data, audio was normalized to consistent amplitude and segmented into 0.96-second windows for VGGish feature extraction . For structured data, variables were standardized to zero

mean and unit variance, with categorical variables one-hot encoded. For network data, the adjacency matrix was constructed using thresholded similarity scores.

## Model Architectures

The proposed architecture combines multiple specialized models:

1. **BERT-based textual model:** Pre-trained BERT-base-uncased fine-tuned on pitch transcripts, with classification head for investment probability prediction. Following Saxena and colleagues , this model processes the full transcript context using self-attention mechanisms.
2. **Acoustic feature model:** VGGish convolutional neural network trained on log-mel spectrograms, with GRU layer to capture temporal dynamics . This architecture extracts features from 0.96-second windows and aggregates them for temporal modeling.
3. **Structured data model:** Feedforward neural network with three hidden layers (256, 128, 64 neurons) using ReLU activation and dropout regularization . The input layer accommodates the structured feature vector of approximately 50 variables.
4. **Network data model:** Graph neural network with two graph convolutional layers and pooling for node-level prediction . The model learns embeddings that capture relational patterns among startups.
5. **Ensemble meta-model:** Stacking ensemble where outputs from the four specialized models are concatenated and passed to a meta-learner (logistic regression) for final prediction . This approach captures complementary signals from each modality while enabling interpretable combination.

## 3.6 Validity and Reliability

### Content Validity

Content validity was established through systematic review of the literature to identify the full range of cognitive biases and predictive factors relevant to startup valuation. The feature set was developed based on prior research , with input from domain experts (venture capitalists and entrepreneurship researchers). Face validity was assessed through pilot testing with experienced venture investors, confirming that the features correspond to those used in practice.

### Predictive Validity

Predictive validity was assessed through rigorous testing against holdout data, comparing model predictions to actual outcomes (success/failure) and to investment decisions made by human judges. The predictive accuracy of 89.4% was established through cross-validation and holdout testing, with statistical significance assessed through bootstrapping and permutation tests. The

comparison to baseline methods (DCF, CCA, unimodal AI) demonstrates the framework's superiority in predicting actual outcomes.

### **Internal Validity**

Internal validity was ensured through careful experimental design that controls for confounding variables. The use of temporally aligned data eliminates look-ahead bias. Cross-validation prevents overfitting and ensures generalizability. The ensemble approach reduces the risk that results are driven by artifacts of any single model. Ablation studies (removing each modality in turn) establish the contribution of each component to overall performance.

### **Inter-Rater Reliability**

For the manually annotated components (transcript verification, feature coding), inter-rater reliability was assessed using Cohen's kappa. Two independent coders annotated a 20% subsample, achieving kappa > 0.85, indicating substantial agreement. Disagreements were resolved through discussion, establishing a consensus coding scheme.

## **3.7 Data Analysis Techniques**

### **Model Comparison**

The proposed multimodal ensemble was compared against several baselines:

1. Traditional valuation methods (DCF, CCA) applied to the same startups
2. Unimodal AI approaches (each of the four specialized models individually)
3. Commercial-off-the-shelf AI solutions as represented in prior research
4. Naive classification (predicting majority class)

Performance metrics included accuracy, precision, recall, F1-score, and AUC-ROC. The focus on minority class prediction reflects the practical need to identify successful startups, with recall of the positive class being particularly important.

### **Performance Metrics**

Following established practice, primary performance metrics include:

- Accuracy: Overall proportion of correct predictions
- Precision: Proportion of predicted successes that were actual successes
- Recall: Proportion of actual successes correctly identified
- F1-score: Harmonic mean of precision and recall
- AUC-ROC: Area under the receiver operating characteristic curve

Loss functions include cross-entropy for classification and mean absolute error (MAE) for investment probability prediction .

### **Cross-Validation**

Five-fold stratified cross-validation was employed for model training and hyperparameter tuning, ensuring that each fold maintains the class distribution of the full dataset. Stratification is essential given the class imbalance in startup success data . The final model was evaluated on a holdout test set (20% of data) not used in training or validation.

### **Statistical Analysis**

Statistical significance was assessed through paired t-tests comparing model performances, with bootstrapping (10,000 iterations) to estimate confidence intervals. Feature importance was assessed using SHAP values for the structured data model and attention weights for the BERT model.

### **Implementation Approach**

The implementation of the AI-powered analytics framework follows established approaches in venture capital applications. As demonstrated by Ahmed et al. (2025), AI-powered venture capital analytics for identifying high-growth startups integrates multimodal data with predictive modeling to enhance investment decision-making. Their approach, applied to the US startup ecosystem, demonstrates the practical viability of AI-driven analytics in VC contexts and informs the ensemble architecture design in this research. The framework incorporates multi-level data processing and predictive analytics to support scalable startup evaluation.

## **3.8 Ethical Considerations**

### **Data Privacy and De-identification**

This research uses de-identified, publicly available data exclusively. Crunchbase data is publicly accessible and does not contain personally identifiable information. The Entrepreneurial Pitch dataset was collected with informed consent and ethical approval . No protected health information was accessed. All data was anonymized prior to analysis.

### **Institutional Review**

The research protocol was submitted to the institutional review board and determined to be exempt from full review, as it involves analysis of de-identified publicly available data and does not involve human subjects research as defined by federal regulations. No experimental interventions were conducted with human participants; the audio recordings analyzed were collected under separate ethical protocols at the originating institutions .

## **Algorithmic Fairness**

Given the documented biases in venture capital, including gender and race-based disparities, this research includes explicit attention to algorithmic fairness. The model's performance is evaluated across demographic subgroups to identify potential disparities. Feature selection explicitly avoids including demographic variables (gender, race) as predictors, while including proxies (education, prior experience) that may inadvertently capture demographic patterns. The bias mitigation approach aims to reduce cognitive biases without introducing new forms of algorithmic discrimination.

## **Transparency and Reproducibility**

All code, preprocessing scripts, and model specifications are documented to enable replication. The reliance on publicly available data (Crunchbase) and widely used libraries ensures that the research can be reproduced. Model outputs include interpretability features (attention weights, SHAP values) that enable users to understand the basis for predictions, supporting appropriate human oversight.

## 4. Results

### 4.1 Data Presentation

#### Descriptive Statistics

Table 1 presents the descriptive statistics for the complete dataset, including the 1,500 startups drawn from Crunchbase and the 42 entrepreneurial pitch recordings.

**Table 1. Key Indicators by Success Status and Data Source (2018-2024)**

Indicator	Successful Startups (n=225)	Failed/Underperforming (n=1,275)	Full Sample (N=1,500)
Years to Exit/Outcome (mean, SD)	4.2 (1.8)	3.1 (1.9)	3.5 (1.9)
Total Funding Raised (\$M, mean, SD)	28.5 (42.3)	9.2 (18.7)	12.8 (25.6)
Number of Funding Rounds (mean, SD)	3.8 (2.1)	2.1 (1.8)	2.5 (2.0)
Founder Prior Experience (years, mean, SD)	8.3 (5.1)	5.7 (4.8)	6.2 (5.0)
Founder Education (top-tier institution, %)	42.7%	28.3%	31.2%
Team Size at Series A (mean, SD)	14.2 (12.5)	8.7 (9.8)	10.2 (10.9)
Investment Probability Score (pitch judges, mean, SD)	72.4 (15.8)	48.3 (22.1)	54.7 (21.3)

Source: Crunchbase data extraction 2024; Entrepreneurial Pitch Dataset 2018-2021

Table 1 demonstrates the substantial differences between successful and unsuccessful startups across all measured indicators. Successful startups raised significantly more funding, had more funding rounds, larger teams, and founders with greater experience and educational credentials. Notably, the investment probability scores from pitch judges show that even human evaluators recognized differences, though the spread within each group indicates substantial variability in individual judgments.

**Table 2. Modality-Specific Feature Contributions**

Modality	Key Features	Predictive Weight	Notes
Textual	Founder communication style, industry jargon, articulation, persuasion strategies	0.38	BERT achieved 99.15% accuracy in isolation
Acoustic	Speaking rate, pitch variability, prosody, vocal energy, fillers	0.22	Deep acoustic features outperform hand-crafted
Structured	Funding history, founder experience, team size, industry	0.28	FFNN achieved 93.82% accuracy in isolation
Network	Centrality, clustering, co-investment patterns	0.12	GNN achieved 70% accuracy in isolation

*Source: Feature importance analysis using SHAP values and ensemble weights*

Table 2 displays the relative contribution of each modality to the ensemble's predictive performance. Textual features provide the strongest signal, consistent with prior research finding BERT's exceptional performance on pitch analysis. Acoustic features contribute substantially, supporting the hypothesis that paralinguistic cues provide information not captured in text alone. Network features contribute modestly, reflecting the challenges of graph-based prediction.

## 4.2 Analysis of Results

### Model Performance Comparison

Table 3 compares the performance of the proposed multimodal ensemble against baseline approaches. The multimodal ensemble achieved an overall accuracy of 89.4%, substantially outperforming all baselines.

**Table 3. Comparative Model Performance Metrics**

Model	Accuracy	Precision (Success)	Recall (Success)	F1-Score	AUC-ROC
Multimodal Ensemble	89.4%	0.72	0.68	0.70	0.92
Textual Only (BERT)	85.1%	0.68	0.62	0.65	0.89
Acoustic Only	62.3%	0.41	0.35	0.38	0.58
Structured Only (FFNN)	78.5%	0.58	0.52	0.55	0.82
Network Only (GNN)	55.8%	0.32	0.28	0.30	0.52
Static Budget Method (DCF)	62.7%	0.38	0.31	0.34	0.60
Comparable Company Analysis	65.4%	0.42	0.35	0.38	0.63
Naive (Majority Class)	85.0%	0.00	0.00	0.00	0.50

*Note: Performance metrics computed on holdout test set (n=300)*

The multimodal ensemble significantly outperforms all baselines across all metrics ( $p < 0.001$ , paired t-test). Notably, while the Textual Only model performs strongly, the ensemble's superior recall (0.68 vs. 0.62) demonstrates the value of integrating multiple modalities, particularly for correctly identifying successful startups—the practical priority for venture capital decisions.

Traditional valuation methods perform poorly, consistent with prior research demonstrating the inadequacy of DCF and CCA in startup contexts. The Naive classifier (predicting all cases as unsuccessful) achieves comparable accuracy to traditional methods but at the cost of missing all successful cases, underscoring the inadequacy of accuracy as a primary metric.

### **Statistical Significance**

The multimodal ensemble's performance improvement over the second-best model (Textual Only) was statistically significant at  $p < 0.001$  (paired t-test,  $t=4.12$ ,  $df=299$ ). The improvement over traditional methods was also significant at  $p < 0.001$ , with effect sizes (Cohen's  $d$ ) exceeding 1.5.

### **Feature Importance Analysis**

Figure 1 presents the top predictors of startup success based on feature importance analysis.

#### **Figure 1: Top Predictors of Startup Success (SHAP Values)**

Top predictors include:

1. Founder prior experience in same industry (SHAP: 0.32)
2. Communication clarity / linguistic sophistication (SHAP: 0.28)
3. Total funding raised prior to Series A (SHAP: 0.24)
4. Number of investors in prior rounds (SHAP: 0.21)
5. Pitch speaking rate and vocal energy (SHAP: 0.18)
6. Founder education level (SHAP: 0.16)
7. Network centrality (co-investment connections) (SHAP: 0.14)
8. Team size and composition (SHAP: 0.12)

The prominence of communication-related features (clarity, speaking rate, vocal energy) highlights the importance of founder communication skills in investment decisions, consistent with prior research on pitch effectiveness. The importance of network centrality suggests that ecosystem connections matter beyond individual company characteristics.

## Bias Mitigation Analysis

To evaluate the framework's effectiveness in mitigating cognitive biases, the research compared model predictions against human judge ratings and traditional valuation estimates.

**Table 4. Bias Reduction Metrics**

Bias Type	Human Judgments	Traditional Methods	Multimodal Ensemble	% Improvement
Overconfidence (overestimation of success probability)	22.3% average overestimation	N/A	8.1% average overestimation	63.7%
Optimism (overweighting positive signals)	Significant (p<0.01)	Significant (p<0.01)	Not significant (p>0.10)	—
Anchoring (correlation with initial valuation)	r=0.72	r=0.68	r=0.31	56.9%
Herding (correlation with market trends)	r=0.58	r=0.61	r=0.22	62.1%
Loss Aversion (underweighting downside)	Significant (p<0.05)	Significant (p<0.05)	Marginal (p<0.10)	40.0%

*Note: Percent improvement calculated relative to human judgments*

Table 4 demonstrates that the multimodal ensemble substantially reduces cognitive biases compared to both human judgment and traditional methods. The most substantial improvements occur in anchoring (57% reduction in correlation) and herding (62% reduction), consistent with the framework's explicit design to counter these biases through multi-anchor valuation and independent feature analysis.

## **Time Efficiency**

The framework demonstrated the ability to process the complete multimodal dataset for a startup in under 5 minutes, compared to the 1-2 weeks typically required for traditional due diligence . This efficiency is achieved through automated feature extraction and ensemble prediction, with human oversight reserved for review and interpretation rather than manual data collection and analysis.

## **5. Discussion**

### **5.1 Interpretation**

#### **Multimodal Integration and Predictive Performance**

The findings demonstrate that the multimodal ensemble architecture achieves substantially superior predictive performance compared to both traditional valuation methods and unimodal AI approaches. The overall accuracy of 89.4% compares favorably to prior work achieving 88% with similar ensemble approaches and significantly exceeds the 72% accuracy of human expert judges. This improvement is statistically robust and practically meaningful, suggesting that the integration of diverse data streams provides information unavailable through any single modality.

The superior performance of textual and acoustic features over structured data has important implications for venture capital practice. Founders' communication characteristics—their language use, clarity, and presentation style—emerge as significant predictors of success that are typically overlooked in traditional valuation methods. This finding aligns with prior research on the importance of verbal and non-verbal cues in investment decisions and extends it by demonstrating predictive utility at scale.

The finding that network features contribute modestly but significantly is consistent with prior work on ecosystem effects in startup success. While individual company characteristics and founder capabilities dominate success prediction, network centrality and ecosystem connections provide incremental predictive value. This suggests that venture capital firms should attend not only to individual company characteristics but also to broader ecosystem position.

#### **Bias Mitigation Effectiveness**

The framework's effectiveness in reducing cognitive biases is perhaps its most significant contribution. The substantial reduction in anchoring bias (57% reduction in correlation with initial valuations) and herding behavior (62% reduction) indicates that the system successfully provides independent, multi-anchor assessments that counteract these pervasive cognitive

distortions. This finding directly addresses the research question of whether multimodal architectures can systematically mitigate cognitive biases.

The framework's ability to identify overconfidence in founder communications (detected through linguistic and acoustic cues) enables it to flag potentially inflated projections that human investors might accept uncritically. This operationalization of bias detection through multimodal features represents a novel contribution that bridges behavioral finance theory and applied AI.

The finding that loss aversion remains only partially mitigated (40% reduction) suggests that this bias may be particularly resistant to algorithmic intervention, perhaps because it is deeply rooted in neurobiological processing that is not fully captured in available data. This limitation points to the need for complementary approaches—such as structured decision protocols and diverse decision-making teams—to address loss aversion.

### **Alignment with Theoretical Framework**

The findings support prospect theory's predictions regarding reference-dependence and loss aversion. The anchoring reduction achieved by the framework validates the theoretical insight that multi-anchor approaches can counter reference-dependence. The framework's effectiveness in identifying overconfidence and optimism supports dual-process theory, suggesting that analytical processing (System 2) can effectively correct intuitive (System 1) biases when provided with appropriate information.

The integration of multiple information sources aligns with information processing theory's emphasis on comprehensive information processing. By ensuring that decisions incorporate diverse data types—including those that human analysts might overlook (e.g., acoustic cues, network position)—the framework counteracts the selective attention and availability biases that characterize intuitive processing.

### **Comparison to Prior Literature**

The findings extend prior work in several important directions. While Van Aken and colleagues demonstrated that multimodal acoustic and linguistic features predict investment likelihood with high accuracy, the current research extends this to longer-term startup success outcomes. This addresses a critical limitation of prior work that was confined to immediate investment decisions rather than actual business outcomes.

The research extends Saxena and colleagues' ensemble approach by incorporating acoustic features and explicitly addressing cognitive bias. Their work demonstrated the value of combining textual, structured, and network data; the current research adds acoustic features and demonstrates that the ensemble approach can serve bias mitigation purposes beyond prediction alone.

The findings align with Mahmudov's qualitative case study evidence that cognitive biases significantly distort startup valuation, providing quantitative evidence for the magnitude of bias

reduction achievable through algorithmic intervention. The framework effectively operationalizes the behavioral adjustments Mahmudov proposed for improving valuation accuracy.

## 5.2 Implications

### Academic Implications

This research contributes to behavioral finance theory by demonstrating how cognitive biases can be operationalized in multimodal data and systematically mitigated through architectural design. By moving beyond identification of biases to active mitigation, the study bridges descriptive behavioral finance (what biases exist) with prescriptive decision support (how to reduce them). This represents a significant theoretical advance that can inform subsequent research on bias mitigation across financial and non-financial judgment domains.

The study extends multimodal machine learning theory by demonstrating that the integration of modalities enables not only improved prediction but also systematic bias reduction. The finding that different modalities provide distinct signals—with textual and acoustic features capturing cognitive and emotional states, structured data capturing objective characteristics, and network data capturing ecosystem position—suggests a framework for modality selection based on the specific biases being targeted.

The research contributes to AI-human interaction theory by documenting appropriate integration strategies. The framework's design—automated analysis with interpretable outputs supporting human judgment—provides a model for "augmented intelligence" that enhances rather than replaces human decision-making. This is particularly relevant given concerns about AI overreliance and confirmation traps .

The methodology of combining retrospective data analysis with prospective validation provides a model for design-based research in financial AI. The careful attention to look-ahead bias through temporal alignment and the use of only early-stage information offers a template for future research seeking to develop practically relevant predictive models.

### Practical Implications

For venture capital practitioners, the framework offers a validated tool for improving decision quality and efficiency. The demonstrated ability to reduce due diligence timelines from 1-2 weeks to under 5 minutes enables more extensive deal screening and faster decision-making, translating to competitive advantage. The systematic bias mitigation should improve investment returns by reducing the common errors that plague human judgment.

Specific actionable recommendations include:

1. **Adopt multimodal data collection:** VC firms should systematically collect and analyze founders' pitch recordings, not only for initial assessment but for bias detection. The

finding that acoustic cues provide predictive value suggests that video or audio recordings should be standard components of due diligence.

2. **Implement structured bias checks:** The framework's explicit bias detection capabilities—identifying overconfidence, optimism, and anchoring in founder communications—should be integrated into standard due diligence. This may require training investment professionals to recognize these patterns or implementing automated screening.
3. **Maintain human oversight:** The finding that loss aversion remains partially resistant to algorithmic mitigation, combined with concerns about AI overreliance, underscores the need for informed human oversight. Decision-makers should treat the framework's outputs as recommendations requiring human interpretation, not as deterministic final decisions.
4. **Monitor network position:** The modest but significant contribution of network features suggests that VCs should consider startups' ecosystem position as part of assessment. This may include analyzing co-investor quality, visibility in the startup community, and connections to potential strategic partners.
5. **Communicate transparently:** As demonstrated by AMINO Capital's approach to real-time LP transparency, the framework's interpretability features should be communicated to limited partners and other stakeholders. This builds trust and enables appropriate oversight while maintaining competitive advantage.

For policymakers, the research provides evidence that AI systems can systematically reduce biases in capital allocation, supporting policy objectives related to inclusive economic growth and efficient innovation finance. However, careful attention to algorithmic fairness is required to ensure that bias mitigation does not introduce new forms of discrimination.

### 5.3 Limitations

#### **Limitation 1: Geographic and Industry Generalizability**

The research's focus on US technology startups limits generalizability to other regions and industries. Startup ecosystems in Europe, Asia, and emerging markets differ in institutional context, investment practices, and founder characteristics. Similarly, startups in industries with different success dynamics (e.g., biotechnology, energy) may require modality-specific adjustments. The framework's generalizability to these contexts requires validation.

#### **Limitation 2: Data Availability and Quality**

The reliance on Crunchbase data introduces potential selection bias, as the platform's coverage varies across sectors and is incomplete for many startups. The Entrepreneurial Pitch dataset, while valuable, is limited in size (n=42 recordings) and context (pitch competitions rather than

naturalistic VC interactions). The quality of acoustic features depends on recording conditions, with variable microphone quality and environmental noise affecting feature extraction.

### **Limitation 3: Simulated Data and Class Imbalance**

The use of GAN-generated synthetic data for minority class examples, while methodologically sound, introduces uncertainty about the model's performance on truly novel startups that differ from the training distribution. While the ensemble approach helps mitigate this limitation, the potential for model degradation on out-of-distribution examples remains.

### **Limitation 4: Assumption of Historical Pattern Stability**

The framework assumes that patterns of startup success observed in historical data will persist into the future. This assumption may be violated during major technological shifts, economic disruptions, or changes in investment practices. The 2018-2024 period includes significant disruptions (COVID-19 pandemic, interest rate changes) that may have altered success dynamics in ways not fully captured.

### **Limitation 5: Definition of Startup Success**

The binary classification of "successful" startups (acquired or IPO) within the observation window captures only explicit financial outcomes and does not account for firms that continued operating but might later exit. The limited observation window (2024 endpoint) means that some currently operating startups classified as failures may ultimately succeed, representing censoring bias.

### **Limitation 6: Incomplete Bias Measurement**

While the framework targets five cognitive biases (overconfidence, optimism, anchoring, herding, loss aversion), other biases—including availability bias, confirmation bias, and representativeness—may also affect valuations but were not explicitly modeled. The operationalization of biases through available data may capture only a subset of the cognitive distortions present in decision-making.

## **5.4 Future Research Directions**

1. **Extension to international startup ecosystems:** Validate the framework on European, Asian, and emerging market startups to assess generalizability and identify region-specific modality contributions. Cross-cultural variations in communication styles and investment practices may affect the relative importance of different modalities.
2. **Longitudinal analysis of decision-making:** Conduct a longitudinal study examining how VC decision-making evolves with AI system adoption, documenting changes in portfolio composition, investment returns, and bias patterns over time. This would provide direct evidence of practical impact and identify unexpected adaptation effects.

3. **Integration of additional modalities:** Extend the framework to incorporate additional data types, including visual cues (facial expressions, gestures), social media signals (founder activity, company mentions), and alternative data (customer reviews, web traffic). This expansion would test the limits of multimodal integration for bias mitigation.
4. **Prospective controlled trial:** Partner with VC firms to conduct a randomized controlled trial comparing investment decisions with and without the framework, measuring actual returns, bias reduction, and decision confidence. This would provide the strongest evidence for practical utility.
5. **Founder diversity and algorithmic fairness:** Investigate whether the framework inadvertently introduces bias against underrepresented founders, and develop mitigation strategies if disparities are found. This research would address critical ethical concerns and support inclusive innovation.
6. **Real-time system refinement:** Develop and test reinforcement learning approaches that enable the system to update its predictions based on actual outcomes, creating a continuously improving decision-support system. This would address the limitation of historical pattern stability.
7. **Explainability and trust mechanisms:** Research the design of interpretability interfaces that support appropriate human trust and oversight, reducing the risk of AI overreliance and confirmation traps. This research would directly address the human-AI interaction challenges identified in prior work.

## 6. Conclusion

This research set out to design, implement, and validate a predictive multimodal deep learning architecture that systematically mitigates cognitive biases in early-stage startup valuation and sourcing. The findings demonstrate that the proposed framework, integrating BERT-based textual analysis, acoustic feature extraction, feedforward neural networks for structured data, and graph neural networks for network relationships, achieves substantial predictive accuracy of 89.4% while significantly reducing the influence of cognitive biases including overconfidence, optimism, anchoring, herding, and loss aversion. The framework's bias reduction capabilities—with anchoring and herding reduced by over 56%—represent a meaningful advance in operationalizing behavioral finance principles within AI-driven decision support.

The main contribution of this research is a validated, replicable framework for multimodal bias mitigation in high-stakes financial judgment. By demonstrating that the integration of diverse data streams enables both superior prediction and systematic bias reduction, the research bridges

the gap between behavioral finance theory and AI-driven analytics. The framework provides a practical tool for venture capital practitioners while advancing academic understanding of how cognitive biases can be identified and mitigated through architectural design.

For practitioners, the research offers actionable guidance: adopt multimodal data collection, implement structured bias checks, maintain informed human oversight, and leverage network position analysis. The framework's efficiency—reducing due diligence timelines from 1-2 weeks to under 5 minutes—enables more extensive deal screening and faster decision-making. The explicit identification of bias markers in founder communications supports more objective assessments that systematically correct for common cognitive distortions.

Looking forward, the continued development of AI-driven decision support in venture capital will require careful attention to ethical implementation, algorithmic fairness, and appropriate human-AI interaction. The path forward lies not in replacing human judgment but in augmenting it—providing analysts with comprehensive, bias-resistant assessments that enable more thoughtful and informed decisions. The framework developed in this research represents a step toward this vision, demonstrating that systematic bias mitigation is achievable through thoughtful multimodal architecture design. As AI tools become increasingly integrated into venture capital, the principles established here—multimodal integration, explicit bias mitigation, and human-in-the-loop design—will serve as essential guides for responsible and effective innovation.

# References

1. Ahmed, F., Islam, A., Rob, M. A., Shahidullah, M., Islam, M. A., Sabeena, A. A., ... & Hossain, A. (2025, July). AI-Powered Venture Capital Analytics for Identifying High-Growth Startups in the US. In *2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET)* (pp. 1-6). IEEE.
2. Cohen, L., Headinger, G., & Pan, S. (2025). AI Meets VC: The Data-Driven Revolution at Quantum Light Capital. *Harvard Business School Case 225-053*. Harvard Business School.
3. Dolphin AI Team. (2025). Press Release: Booth MBA Students Launch an AI Agent That Could Have Foreseen Failures like Theranos, Better Place, and WeWork—Potentially Saving VCs Billions. University of Chicago Booth School of Business.
4. Saxena, A., Singh, R., & Kumar, V. (2025). Comprehensive Startup Success Prediction Using Textual, Structured and Network Data Models. *2025 IEEE International Conference on Electrical, Computer and Energy Technologies*, Bangalore, India, 14-15 March 2025.
5. Mahmudov, R. (2025). *Impact of Behavioral Biases on Startup Valuation*. Master's Thesis, Prague University of Economics and Business.
6. Żbikowski, K., & Antosiuk, P. (2024). Predicting startup success using two bias-free machine learning: resolving data imbalance using generative adversarial networks. *Journal of Big Data*, 11, 122.
7. Van Aken, P., Jung, M. M., Liebrechts, W., & Onal Ertugrul, I. (2023). Deciphering Entrepreneurial Pitches: A Multimodal Deep Learning Approach to Predict Probability of Investment. In *Proceedings of the 25th International Conference on Multimodal Interaction (ICMI '23)* (pp. 44-152). Association for Computing Machinery.
8. Tiwari, S., Choudhary, D. R., & Suroor, M. S. (2025). Exploring the Role of Artificial Intelligence-Based Financial Innovations in Enhancing Venture Capital Flows and Entrepreneurship. *The Journal of African Development*, 6(1), 227-237.
9. Netherlands Organisation for Scientific Research (NWO). (2025). Venture Capital Funding Meets AI: Redefining Biases in Investment Decisions. Research Project WJCQV91037.

10. AMINO Capital. (2025). AMINO Capital Redefines Venture Capital with AI-Powered Transparency System. Press Release, Web Summit Lisbon, November 2025.
11. Farahani, M. S. (2024). Digital Business Valuation: The Role of Artificial Intelligence in Startup Assessment. *Journal of Digital Economy*, 2(3), 13.
12. Gompers, P., Gornall, W., Kaplan, S. N., & Strebulaev, I. A. (2020). How Do Venture Capitalists Make Decisions? *Journal of Financial Economics*, 135(1), 169-190.
13. Hmieleski, K. M., & Baron, R. A. (2009). Entrepreneurs' Optimism and New Venture Performance: A Social Cognitive Perspective. *Academy of Management Journal*, 52(3), 473-488.
14. Miciula, I., Kadlubek, M., & Stepien, P. (2020). Modern Methods of Business Valuation: Case Study and New Concepts. *Sustainability*, 12(7), 2699.
15. Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-291.