《当代教学研究学报》
2026 年 11 月，第 3 期: 7—17.　DOI:10.66208/j.jctr0868.6.3.5

AI 动作识别赋能高校武术精准教学的实践路径与效果研究

牛广

（镇江市高等专科学校 体育部，江苏 镇江 212000）

摘要：高校武术普修课班额大、动作瞬时完成，教师凭肉眼难以逐人纠错，技术评价长期依赖主观印象。研究把动作识别定位为“形的测量仪”，而不是“功夫的裁判”，构建采集、识别、诊断、决策、反馈五层结构，以诊断、处方、复测三者的闭合程度界定教学精准。在某高职院校体育教育专业两个普修班（117 人）开展 16 周准实验：实验班技能测评由 68.4 分升至 86.1 分，对照班由 68.1 分升至 78.9 分，后测组间差异有统计学意义（t=5.57，P<0.001，d=1.03，95%CI 0.66~1.44）；达标率 89.8%对 72.4%，课堂练习密度与课外自主练习时长同步上升；技能自我效能提高，理论测验未见差异。识别模块在 5 个动作上的分类准确率为 90.2%，关节角平均绝对误差 4.2°，与教师评分的相关系数为 0.84。研究认为，动作识别的作用集中在诊断与反馈两个环节，其教学价值取决于处方是否可行、反馈是否撤得掉，而不取决于算法精度本身。

关键词：AI 动作识别；精准教学；高校武术；反馈设计；准实验

AI-Based Motion Recognition for Precision Wushu Teaching in Higher Education: Implementation Path and Effects

Niu Guang

(Department of Physical Education, Zhenjiang College, Zhenjiang 212000, China)

Abstract: Large classes and fast-moving techniques make it hard for teachers of college Wushu courses to correct every student by eye, and skill assessment has long rested on subjective impression. This study treats motion recognition as an instrument for measuring form rather than a judge of martial proficiency, and builds a five-layer structure of capture, recognition, diagnosis, decision and feedback, in which precision is defined by how tightly diagnosis, prescription and re-testing are linked. A 16-week quasi-experiment was conducted in two Wushu classes (n=117) at a vocational college. The experimental class rose from 68.4 to 86.1 points and the control class from 68.1 to 78.9; the post-test difference was significant (t=5.57, P<0.001, d=1.03, 95% CI 0.66 to 1.44). The pass rate was 89.8% versus 72.4%, and practice density and out-of-class practice time increased in parallel. Skill self-efficacy improved, whereas the written theory test showed no difference. The recognition module reached 90.2% classification accuracy across five techniques, with a mean absolute joint-angle error of 4.2° and a correlation of 0.84 with teacher ratings. The study concludes that motion recognition works mainly at the stages of diagnosis and feedback, and that its instructional value depends on whether prescriptions are workable and whether feedback can be withdrawn, not on algorithmic accuracy alone.

Key words: AI-based motion recognition; precision teaching; college Wushu; feedback design; quasi-experiment

收稿日期：2026-10-05；在线发表日期：2026-11-10.
第一作者：牛广（1977—），男，硕士，主要从事学校体育教学研究.

一间武术房，一块垫子，五十八名学生，四十分钟。教师示范“弓步冲拳接弹踢”，学生跟着做，整套动作二十秒完成。教师从第一排看到最后一排，看完时第一排已经忘了自己哪里不对。这不是不负责，是肉眼的分辨率有限：一个人同时只能盯住五六个人，而武术要看的部位从脚尖到眼神有十几处。

武术教学长期依赖“看”。手型、步型、发力顺序、重心高低都在教师眼里，学生得到的反馈多半是“再来一遍”或“后腿蹬直”。麻烦在于，这句话对不同的人意味着不同的动作，而教师没有时间逐个确认。

教育数字化被期待把“看不见”变成“看得见”。董倩与刘阳把学校体育数字化分为起步、应用、融合、转型四个阶段，指出四类梗阻：教师数字素养不强、产品与教学融合困难、数据挖掘不深、协同机制不通[1]。于素梅与王晓燕则强调，落点不是技术本身，而是人机协同下教学精准化与育人全面化的统一[2]。方向清楚，剩下的是操作：数据从哪儿来，来了给谁看，看完改什么。

人工智能进入体育学习，已被描述为从“经验判断”走向“智能纠偏”[3]。这话有道理，但要打折扣。郭江浩与熊文从技术现象学提醒，技术对学校体育的赋能是有边界的，身体的、情境的、关系的东西未必能被传感器捕捉[4]。两边都成立：技术看得见人看不见的，也看不见人最想教的。

武术研究对技术的兴趣来得不晚。曾天雪等分析长兵“攻击前手”的技术，把难以言传的发力拆成可比较的参数与图形[5]，证明技术细节可以被测量，但落点在描述而非课堂。课堂这一端，刘文武与王梦东的诊断更贴近日常：组织形式单一封闭，缺少合作互动，学生难以获得效能感[6]。两端各有进展，中间那段路没人走。

动作识别提供了走这段路的可能。Pang 等对武术骨架序列做特征对齐，用多模型加权决策给出质量评分，接近专家水平，并能指出扣分来自哪个环节[7]。Sianto 与 Kristian 用 BlazePose 提取关键点，用动态时间规整把学生动作与标准视频对齐，在 139 段视频上测试，二维角度模型的平均准确率为 82.9%，并能定位出错部位[8]。一个偏重打分，一个偏重纠错。

但实验室的准确率不等于课堂的可用性。Yuan 等在自建太极拳数据集上做细粒度识别，发现小样本下的性能仍受制于数据规模与关节匹配[9]。Hamilton 等把 MediaPipe、HRNet 与三维动作捕捉逐项比对，关节角变异系数多在 10% 以内，屈伸类动作的一致性优于全身性动作[10]。译成课堂语言：系统看得准“胳膊弯了几度”，看不准“这一下有没有沉下去”。

还有一层常被忽略：反馈有剂量。Weakley 等对 20 项阻力训练研究做元分析，视觉反馈对动作速度的效应量为 Hedges’ g=1.11，言语反馈为 g=0.47；频率越高即时效果越好，一旦撤除，效果立刻回到基线[11]。给得越多不等于学得越好，撤得对不对才是关键。

研究要回答三个问题：识别输出怎样转成教学决策；转成之后学生的技能、练习行为与自我评价会不会变；技术误差在武术教学里会被放大还是被吸收。前一个靠路径建构回答，后两个靠 16 周准实验回答。

1　技术前提与概念界定

1.1　动作识别在武术课堂能看见什么

一套动作打完，系统能给出的东西不多：每帧 33 个关键点坐标[8]，由此算出的关节角、角速度、重心投影，以及动作分段的时点。把这些量排成序列，就有了“形”的记录。记录的价值在可比较：这个学生这周与上周比，与模板比，误差落在哪个关节、哪段时间，都能指出来。

“形”之外的部分，系统给不了。武术讲的“劲”，是用力与呼吸、意念配合的结果，外面表现为速度变化与整体松紧，用单目视觉反推误差很大。对抗中的时机更麻烦：什么时候进、什么时候撤取决于对手，而对手不在画面里。所以本文的定位是“形的测量仪”，不是“功夫的裁判”。系统只回答“哪里不一样”，不回答“好不好”。这句话要留一个注脚：阈值本身是教师对动作质量的判断，系统执行的还是这套判断，只是执行得更一致、更不知疲倦。所谓不做好坏判断，指的是系统不替代教师裁决，不是它价值中立。

1.2　三类误差及其教学后果

第一类是遮挡。学生间距不足一米，正面机位拍到的常是重叠轮廓，单目又缺深度，前后站立时系

统会把后面那人的手臂算到前面那人身上。处理办法是改错位站列，并丢弃置信度低于 0.5 的帧段，代价是有效数据减少约两成。

第二类是域偏移。公开数据集的样本以日常动作、舞蹈、健身为主[9]，武术的快慢相间、拧转、低姿态很少见。模型不是没见过人，是没见过这样动的人。用 3200 段校内课堂视频微调后，五个动作的准确率提高了 6 到 9 个百分点，低姿态动作的漏检仍未解决。第三类是服装与器械。宽松练功服遮住肘膝，长器械与手臂关键点混淆，误差由此而来。本研究只做拳术、要求穿合身上衣，这是为可行性作的妥协。

三类误差叠加，单条提示的可信度大约八成。这数字不坏，但教学的逻辑与检测不同：漏报只是少提醒一次，误报会让学生把对的改错。一次“膝盖太前”的错误提示，足以让一个本来做得对的学生改上两周。系统的原则因此是宁可漏报，不可误报。

1.3 两种“精准”

测量精准与教学精准不是一回事。前者指仪器误差小，后者指教学决定对路。系统可以把关节角误差控制在 3 度以内，却仍然对课堂无用，因为它给的是一份 6 项指标、每项都带小数的报告，而教师需要的是“这三个人今天练什么”。反过来，精度 85% 的系统若能把全班问题归成三类并给出可执行任务，反而更接近精准。

本文把教学精准定义为诊断、处方、复测三者的闭合程度：能不能在有限时间找出最该改的那一处，能不能给出这个学生做得来的改法，能不能下次课确认改了没有。三者缺一，精准就退化成报表。操作上，闭合落在三个可查的痕迹：同一偏差是否在复测中再现，处方是否留下执行记录，复测是否沿用初诊那套指标。三样都在，才算闭合一次。这套界定仍是定性的，没有折成单一指数，是概念落地的一个缺口。这也解释了后文为何花大量篇幅谈反馈规则，而不是继续优化模型。

2 实践路径

2.1 系统结构

系统分五层：采集、识别、诊断、决策、反馈（图 1）。分层不是为了好看，是让每层可以单独替换：学校换摄像头不影响诊断规则，教研组改阈值不影响识别模型。推广时这一点很实际，多数学校的设备是分批采购的。

采集层用两台固定机位，一正一侧 45 度，自练时用手机补拍。没有采用多目阵列，因为教室不是实验室：布线、标定、场地占用，任何一项都足以让方案落不了地。两台机器加一台边缘服务器约 1.2 万元，普通院系能承受。

识别层用 BlazePose 提取关键点[8]，先以髋宽为尺度、支撑脚为原点做归一化，消除身高与站位影响，再分段。分段没有用端到端时序网络，而是用膝踝速度阈值切出动作单元，再用动态时间规整与模板对齐。好处是可解释：教师能看懂“这一段为什么被判慢”，而不是面对一个 0.73 的置信度发呆。

诊断层把对齐后的序列折算成六项指标，逐项比对阈值，输出偏差的方向与幅度。决策层最不技术，它把个体偏差聚成三类处方：三人以上共有的进课堂集体提示，三到五人进小组任务，个别的进加练名单。反馈层有投影、课后报告、教师看板三条通道。五层之外还有一条回流线，反馈数据回到采集层，构成下一轮复测。没有它，系统就是一台一次性体检仪；有了它，才谈得上教与学的循环。

2.2 三阶段教学流程

流程按课前、课中、课后三段组织（图 2）。课前学生拍一段 10 秒组合动作上传，系统给个人基线，教师在看板上选定当堂要盯的两个指标。这里有一条限制：一节课只盯两个。指标多了，提示互相淹没，学生也不知道先改哪个。

课中分三步。示范后投影打出全班共性偏差前三条，教师据此讲要点，讲解控制在三分钟内；随后按诊断分组，A 组补下肢蹬伸，B 组补节拍，C 组做提高任务，任务卡写明练什么、几组、什么算过关；最后二十分钟按看板名单抽检，一组一组过，不再全班统一纠错。

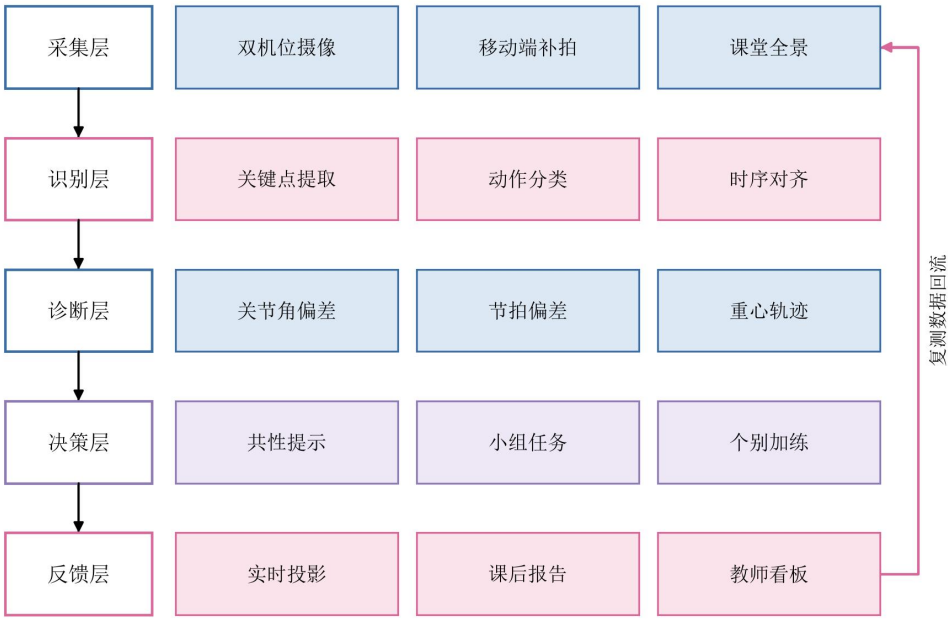


图 1　AI 动作识别辅助武术教学的系统结构

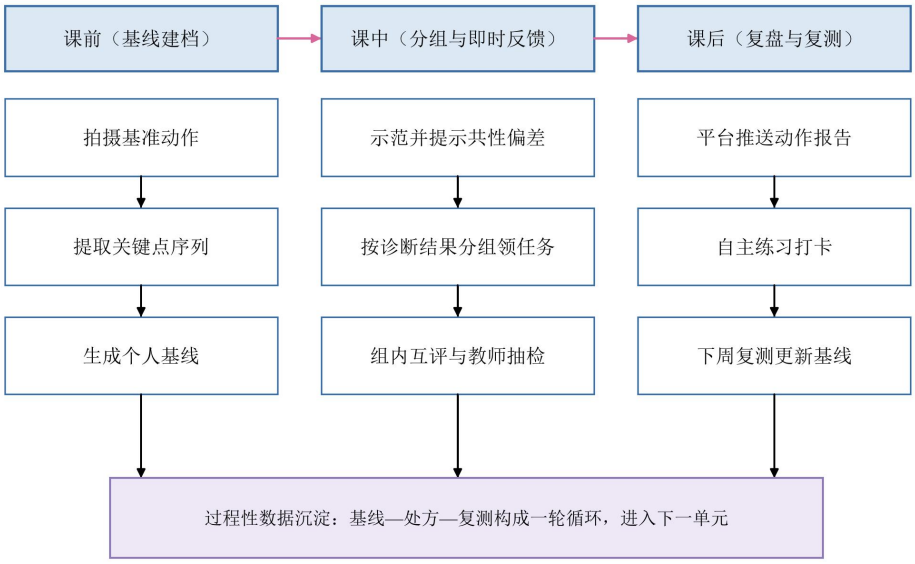


图 2　基于动作识别的三阶段教学流程

课后平台推送当日报告，含关节角曲线、问题片段与两条建议；学生按任务卡自练并打卡；下周课前复测，基线随之更新。三段不是线性流程，而是螺旋：这一轮的复测就是下一轮的基线。

一个细节值得说：课中的投影屏放在场地一角，不在正前方。学生要看，得走过去。这看似降低便利，实际必要：屏幕正对队伍，学生会盯着屏幕做动作，眼睛离开自己的手脚，而武术恰恰要求目光随手走。

2.3　诊断指标与阈值

六项指标与阈值见表 1。阈值不是算法学出来的，是教研组定的：10 名武术专项教师对 200 条学生视频逐条判定，取多数意见，再与教材规范对照折算。费时，但解释权留在教师手里：系统提示时，教师可以追问“凭什么”，答案是一句人话，不是一组权重。

定阈值有一条原则：宁可宽一点。弓步前腿膝角，教材要求大腿接近水平，但髋关节柔韧性差的学

生很难做到，硬按 90 度卡，一半人被判错。最终取 95 度，把“明显不够”与“稍差一点”分开，后者不提示。阈值不是技术参数，是教学判断的量化表达。

指标之间会打架。躯干前倾超过 25 度，可能是塌腰，也可能是低姿态的必要代偿。系统只报数值，定性留给教师。实践中，两条以上同时越界才判定需要干预，单条多半放过。

表 1　六项诊断指标与判定阈值

指标	计算方式	判定阈值	教学含义
前腿膝角	髋—膝—踝夹角	弓步小于 95°	重心前移不足或过度
后腿蹬伸角速度	膝关节角速度峰值	低于 180°/s	蹬地不充分，力未起于腿
躯干前倾角	肩髋连线与铅垂线夹角	大于 25°	塌腰或前俯代偿
支撑脚外展角	足轴线与正前方夹角	大于 30°	步型散，发力方向偏移
动作节拍比	各段耗时与模板之比	偏离 0.85~1.15	快慢相间被拉平
重心左右偏移	髋中心水平位移	大于 0.08 倍身高	单腿支撑不稳

2.4　反馈的投放规则

反馈有三种形态（表 2）：即时视觉、课后报告、教师看板。三者的内容、时点、频率都不同，混在一起就会失效。常见失误是把所有信息塞进即时反馈，学生一边做一边看屏，动作更慢、更犹豫。

投放规则四条：单节课即时提示不超过三条；同一指标连续提示三次后撤除；周报告只在周一推送；看板只显示需干预的学生，不显示排名。第三条防信息疲劳，第四条防评价变味。

撤除直接来自 Weakley 等的结论[11]：视觉反馈即时效应大，撤除后效应立即消失，说明它更像“带练”而不是“学会”。既然依赖是常态，教学设计就必须安排撤除。本研究从第三个月起把即时提示减到一条，第五个月起课堂上不再投影。

不显示排名，是因为排名会把注意力从动作转到名次。系统记录每人的达标进度，只对本人和教师可见。这条规则起初遭质疑，有教师认为“没有比较就没有动力”；一学期后质疑少了，因为比较已由小组任务承担。

表 2　三类反馈的投放设计

形态	内容	投放时点	频率上限	面向对象
即时视觉	关键帧对照、偏差方向	练习间隙	每课 3 条	学生本人
课后报告	关节角曲线、问题片段、两条建议	当日 20:00 前	每周 1 次	学生本人
教师看板	小组热图、个别关注名单	课前 10 分钟	每课 1 次	任课教师
教研汇总	班级指标分布、教学建议	单元结束	每 4 周 1 次	教研组

2.5　教师的位置

系统把 58 人压缩成 6 个小组和 5 个名字。教师进场地前看一眼看板，就知道今天该找谁，不必从队首走到队尾挨个扫。注意力从平均分配转为按需分配，有教师形容，以前是撒胡椒面，现在是点名。

有一条线不能越：提示说不说、怎么说，仍由教师定。算法不知道这个学生昨天崴了脚，不知道那个学生正在准备校运会，更不知道有人一被当众纠正就紧张。系统给候选，教师做决定。这条线守不住，技术就从助手变成干扰。

3　研究设计

3.1　研究对象

研究在某高职院校体育教育专业 2023 级武术普修课开展，取两个平行班：实验班 59 人（男 31、女 28），对照班 58 人（男 30、女 28），平均年龄 19.6±0.8 岁。同一位教师任教，大纲、课时、场地器材一致，入学术科成绩与前一学期体育成绩无显著差异。

选普修课而非专项课，是因为普修课课时少、班额大，教师没有时间精细指导；专项课的学生本来就会自己看视频抠动作，技术干预的空间反而小。但要说清楚：体育教育专业的学生有运动基础，模仿

与术语理解都强于公共体育课学生，本研究的样本因此偏在好教的一端。同一套方案搬到公体课，班额更大、基础更薄，效果大概率要打折，这需要另做验证。

研究经学校教学指导委员会同意，向学生说明目的与流程，签署知情同意书，可随时退出。视频只在校内服务器存储，面部模糊处理，报告不署名，仅以学号后四位标识。

3.2 干预方案

全程 16 周、32 学时，每周 2 学时。内容为初级长拳第三路前两段与 24 式太极拳选段，两班相同。实验班按第 2 节的结构组织，对照班按“示范—模仿—纠错—考核”组织，同样用手机录像，但只在课后抽查，不做关键点分析，也不推送报告。

对照班也录像，是为了排除“被拍摄”本身的影响。准实验常见的麻烦是实验班因被观察而表现更好，若对照班完全不被拍摄，这种差异就会被误读为模式的效果。

教师培训在实验前进行，3 次、6 小时，内容为看板使用、处方设计、提示话术。花时间最多的是话术：同一句“膝角偏大”，说成“前腿再收一点”学生能改，说成“膝关节角度超出阈值 12 度”学生就懵了。技术语言转成教学语言，是最容易被低估的环节。

3.3 测量工具

技能测评用双机位录像、3 名非任课教师盲评，满分 100。评分表含动作规格 40 分、劲力与节奏 25 分、眼神与身法 20 分、完整连贯 15 分，ICC(2,3)=0.86，分歧超过 8 分由第三人裁定。

达标以 80 分为线，依据是该校普修课现行技能档与教研组议定，不是统计中位数。这条线在该校属“良好”，比及格严，取严是想让达标对应可上场演示的水平，代价是达标率整体偏低；换一条线两班的差距会变，这一点留到局限里说。练习密度用加速度计与课堂观察双记录，取单元内 4 节课均值。课外练习时长取平台打卡日志，以实际上传视频时长计，没有第三方校验，代拍与刷时长的可能排不掉，只能靠内容抽查作有限约束。技能自我效能量表 8 题、5 点计分，α 为 0.88。理论测验闭卷 50 题，覆盖武术史、动作要领与裁判规则。

3.4 识别模块的效度检验

从两班课堂视频抽取 1448 段片段，覆盖五个教学内容，由两名专项教师逐帧标注起止时点与关键关节角，标注者一致性 ICC(2,2)=0.91。以人工标注为参照，计算分类准确率、关节角平均绝对误差，以及系统相似度得分与教师评分的相关系数。

没有用三维动作捕捉做参照，是本研究的一个短板。捕捉实验室离教学场地四十公里，把学生拉过去测，测出来的也不是课堂里的动作。用人工标注替代，精度上限受标注者限制，解释结果时需要记住。

3.5 统计分析

使用 SPSS 27.0。组间比较用独立样本 t 检验，组内用配对 t 检验，三个时点用 2（组别）×3（时点）重复测量方差分析，计数资料用 χ^2 检验。效应量统一报告：组间比较用 Cohen’s d，方差分析用偏 eta 平方（η^2_p，partial eta squared）；引用他人研究时保留原文指标并注明类型（如 Hedges’ g）。d、率差与 OR 均附 95%CI，格式统一为“95%CI 下限~上限”。达标率同时报告未校正值、连续性校正值与 Fisher 精确检验。显著性水平 α=0.05；交互效应显著后作简单效应分析，多重比较用 Bonferroni 法校正。

4 结果

4.1 识别模块的精度

五个动作的整体分类准确率 90.2%，关节角平均绝对误差 4.2°，相似度得分与教师评分的相关系数 0.84（P<0.001）。分动作看差距明显（表 3）：弓步冲拳最高，94.5%；提膝穿掌最低，85.1%。低姿态、单腿支撑的动作遮挡最重，误差也最大，与 1.2 节的判断一致。

90.2% 做自动评分不够，做偏差提示够用。评分要求绝对准确，差 3 分就影响结论；提示只要求方向正确，即便角度偏了两度，学生改的方向仍然对。研究因此没有继续追精度，把力气花在阈值与处

方的可行性上。

表 3　识别模块在五个动作上的精度

动作	片段数	分类准确率/%	关节角 MAE/(°)	与教师评分 r
弓步冲拳	328	94.5	3.2	0.89
弹踢推掌	301	92.4	4.1	0.86
马步架打	286	90.2	3.8	0.84
歇步盖打	274	87.6	4.9	0.81
提膝穿掌	259	85.1	5.3	0.78
合计/均值	1 448	90.2	4.2	0.84

注：关节角 MAE 为平均绝对误差；r 为系统相似度得分与 3 名教师评分均值的 Pearson 相关系数，均 $P<0.001$。

4.2　技能成绩

前测两班持平（68.4±7.9 对 68.1±8.2，$t=0.20$，$P=0.841$，$d=0.04$，95%CI -0.33~0.40）。中测实验班领先（78.2±7.1 对 72.6±7.8，$t=4.06$，$P<0.001$，$d=0.75$，95%CI 0.39~1.14），后测差距扩大（86.1±6.3 对 78.9±7.6，$t=5.57$，$P<0.001$，$d=1.03$，95%CI 0.66~1.44）。2（组别）×3（时点）重复测量方差分析显示：时间主效应 $F(2,230)=136.4$，$P<0.001$，$\eta^2_p=0.54$；组别主效应 $F(1,115)=11.8$，$P<0.001$，$\eta^2_p=0.09$；组别×时点交互效应 $F(2,230)=21.7$，$P<0.001$，$\eta^2_p=0.16$。因交互效应显著，进一步作简单效应分析（以混合设计的合并误差项为分母）：时间因素在实验班内的简单效应 $F(2,230)=99.7$，$P<0.001$，$\eta^2_p=0.46$，在对照班内 $F(2,230)=36.7$，$P<0.001$，$\eta^2_p=0.24$，两班前测—中测、中测—后测的两两比较均有统计学意义（$P<0.001$）；组别因素在各时点的简单效应即上述三次组间比较，经 Bonferroni 校正（$\alpha'=0.017$）后，中测与后测差异仍显著，前测无差异。

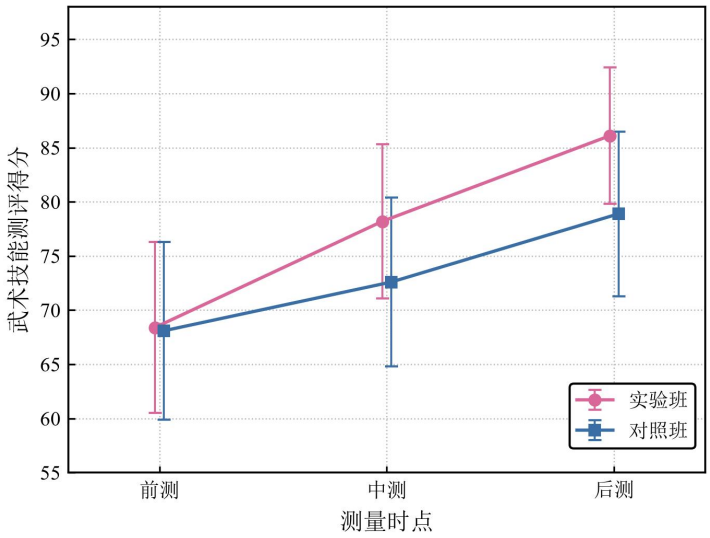


图 3　两班武术技能测评得分的变化（误差棒为标准差）

增量分布比终点差异更能说明问题（图 3）。实验班 16 周净增 17.7 分，对照班 10.8 分。差距主要产生在前 8 周（5.6 分），后 8 周只有 1.6 分。前段差距大，与处方集中在下肢蹬伸和节拍有关，这两项恰是肉眼最难同时盯住、系统最容易量化的；后段收窄，是因为对照班进入考前强化，常规教学的优势显出来了。

4.3　达标率与练习行为

以 80 分为线，实验班达标 53 人（89.8%），对照班 42 人（72.4%），率差 17.4 个百分点（95%CI 3.6~31.3）。未校正 $\chi^2=5.81$、$P=0.016$；连续性校正 $\chi^2=4.73$、$P=0.030$；Fisher 精确检验 $P=0.019$，OR=3.37

（95%CI 1.21~9.35）。三种算法结论一致，与前人研究中常见的“校正后不显著”不同，原因大概是样本量与率差都够。

练习密度由 47.3% 升至 63.8%，对照班由 46.9% 升至 52.1%（t=10.12，P<0.001，d=1.87，95%CI 1.47~2.35）。课外自主练习由 28.6 分钟升至 64.2 分钟，对照班由 29.1 升至 41.5（t=6.91，P<0.001，d=1.28，95%CI 0.90~1.70）。两条曲线在中测后分开（图 4），与技能得分的变化节奏大致同步。

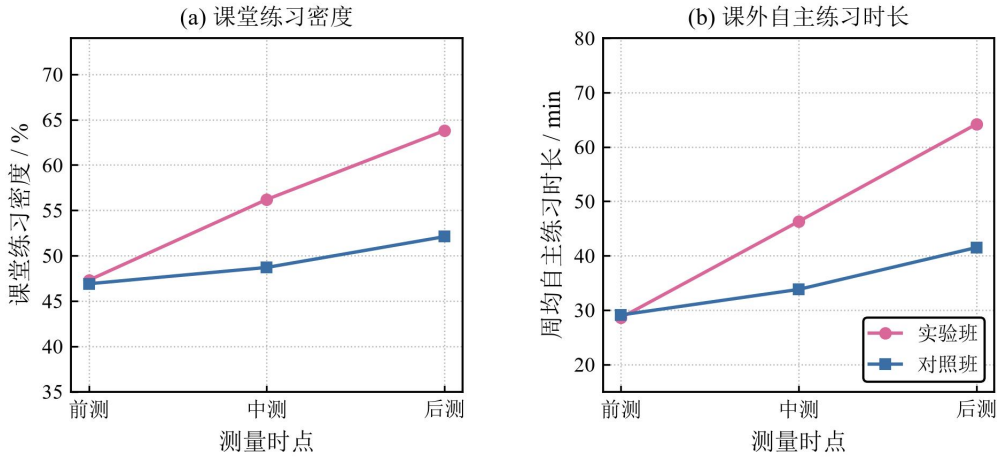


图 4　课堂练习密度与课外自主练习时长的变化

练习密度的效应量偏大，需要谨慎。分组轮换、任务卡、按名单抽检，这些组织改动本身就会提高密度，不能全记在技术头上。比较老实的说法是：技术让分组有了依据，分组提高了密度，密度又支撑了练习量，三者串联，拆不开。

4.4　自我效能与理论成绩

技能自我效能由 3.24 升至 3.91，对照班由 3.21 升至 3.58（t=3.24，P=0.002，d=0.60，95%CI 0.24~0.99）。增幅最大的是“我能看出自己动作的问题”，实验班提高 0.86，对照班 0.31。这与系统把问题指出来、并留下可回看的报告有关：学生不必只等教师开口。这一条可看作自我错误觉察的粗略指标，它的增幅最大，与撤除后成绩没有塌下去也互为印证。不过量表里只有一个条目指向它，中介效应检验没有做，说它解释了撤除效果，证据还不足。

理论测验没有差异（76.4±8.2 对 75.8±8.6，t=0.39，P=0.700，d=0.07，95%CI -0.29~0.44）。这不意外，动作识别作用于动作，不作用于陈述性知识。

4.5　教师工作量

单节课个别指导人次，实验班 38.6，对照班 16.4，翻了一倍多；教师讲授时间反而从 9.8 分钟降到 4.2 分钟，因为共性问题已在投影上说过。一升一降，说明注意力被重新分配，不是简单加量。

总工作量仍增加。实验初期教师每周约 1.5 小时看报告、调分组，一学期后降到 0.6 小时，因为处方模板积累起来。这条下降曲线值得注意：只投入一学期的学校，看到的会是“更累”，容易半途而废。

表 4　两班各指标的前后比较

指标	实验班前测	实验班后测	对照班前测	对照班后测	t	P	d	95%CI
技能测评/分	68.4±7.9	86.1±6.3	68.1±8.2	78.9±7.6	5.57	<0.001	1.03	0.66~1.44
练习密度/%	47.3±6.8	63.8±6.4	46.9±7.1	52.1±6.1	10.12	<0.001	1.87	1.47~2.35
自主练习/min	28.6±15.2	64.2±18.6	29.1±14.8	41.5±16.9	6.91	<0.001	1.28	0.90~1.70
自我效能/分	3.24±0.58	3.91±0.52	3.21±0.61	3.58±0.58	3.24	0.002	0.60	0.24~0.99
理论测验/分	74.2±9.1	76.4±8.2	73.9±8.8	75.8±8.6	0.39	0.700	0.07	-0.29~0.44

注：数值为 M±SD；t、P、d 与 95%CI 为后测组间比较结果；实验班 n=59，对照班 n=58。

5 讨论

5.1 精准的落点在决策

本研究的识别精度只有 90.2%，低于文献中常见的高精度报告[7,9]，但技能成绩的效应量达到 Cohen's d=1.03。这个反差说明，教学效果不取决于识别有多准，而取决于结果有没有变成教师和学生能执行的动作。Pang 等的自动评分接近专家水平[7]，可评分本身不改变教学；Sianto 与 Kristian 的实时纠错准确率 82.9%[8]，但没有报告学习效果。本研究补的正是这一段。

另一面也需提到。体育智能裁判的讨论已经指出，算法进入评分会带来技术风险与价值风险，评价权从教师部分转移到系统[12]。课堂虽不至于如此严重，但“谁的判断算数”这个问题，系统上线第一天就出现了。

这也解释了一个常见现象：不少学校买了设备，两个月就搁置。设备没坏，是数据没有出口。报告发给谁、谁看、看完改什么，这三件事没有答案，再准的算法也只是多了一块屏。

5.2 反馈的剂量与依赖

即时反馈越多越好，是直觉，不是事实。Weakley 等的元分析显示，视觉反馈的即时效应量达 Hedges' g=1.11，撤除后效应立即消失[11]，这与运动学习里的指导假说一致：外部提示替代了学习者自身的错误检测，提示一撤，检测能力没留下[13]。Wulf 等在综述里把反馈频率与注意焦点并列，认为过密的外部反馈会让学习者盯着提示做，而不再盯着自己的身体。

本研究因此把撤除写进流程。撤除后实验班成绩没有回落：后 8 周两班差距虽收窄到 1.6 分，实验班的绝对得分仍在上升，说明减量设计起了作用。但有 8 人出现小幅回落，回看记录，共同点是课外练习低于 30 分钟。提示可以撤，练习量撤不得，这一点只能靠任务卡与打卡来兜。还有两处没做干净。减量按时间统一执行，没有按达标进度分层，基础好的本可以更早撤，基础弱的或许还要多带两周，这个节奏没有试。小组任务里的互评与相互模仿也在起作用，分组本身就带来观察和纠正，本研究没有把它从反馈效应里拆出来。

5.3 误差在教学里的放大与吸收

4.2° 的平均误差在工程上算不错[10]，在武术教学里却可能被放大。弓步膝角差 4 度，对应的正是“膝尖过不过脚尖”这个最常纠的判断，误差一旦跨过阈值，提示就指错方向。本研究设了两道闸：置信度低于 0.5 的帧段不生成提示；同一提示连续三帧越界才推送。人工复核 420 条提示，误报 24 条，误报率 5.7%。

误差也有被吸收的一面。武术的很多要求本来就是区间：重心高低、步幅大小、手臂弧度，都允许个人差异。阈值定宽一点，误差反而被吸收进允许范围。这与工程上追求最小误差相反，但教学要的是可用判断，不是精确读数。再往下还有一笔账，本研究算不了，却不能不说。武术要传的不只是关节角，还有劲力、身法与精气神，这些恰是传感器最难接住的部分，也正是郭江浩与熊文所说技术限度的一部分[4]。资源一旦向可量化的指标倾斜，课堂时间与教师的注意力就跟着走，不可量化的内容容易被挤到边上。测量仪越顺手，越要留一块不测量的地方。

5.4 教师被重新分配，而不是被替代

实验里最明显的变化不是学生，是教师。以前教师站队前扫全场，靠经验挑典型；现在进场地前看看板，直接走到该找的人面前。注意力总量没变，分配方式变了。这与董倩等指出的“数据挖掘不深、决策支持不强”正好对上[1]：数据的价值不在多，在于能不能支持一个具体决定。

但也带来新负担。教师需要读懂关节角曲线，判断提示该不该说，在一堆数字里挑出那一条。6 小时培训能解决操作，“什么时候不该信系统”教不会，只能一学期一学期积累。设备可以采购，判断力不能。

5.5 与既有研究的对话

把本研究放回文献，可以看到三条线索：教育数字化的宏观判断，强调人机协同与循证决策[1-2]；

人工智能进入体育学习的样态描述，强调从经验判断到智能纠偏[3]，同时提醒技术限度[4]；武术领域的技术尝试，从参数化[5]到自动评分[7]再到实时纠错[8]。三条各自成立，缺的是把技术输出接到课堂决策上的中间层。本文的五层结构与三阶段流程，是这个中间层的一个版本。与国内同类干预比，差别不在技术新，而在把口径摊开：技能、达标率、练习行为、自我效能、理论成绩、教师工作量一并报告，没效的也说。不少数字化教学实验停在成绩提高、兴趣上升两句结论上，读者看不出效果来自技术还是来自加课加练。

这个版本有明显的地方性。它依赖两台机位、一台边缘服务器、一名会用看板的教师、一个愿意定阈值的教研组。换到设备更差、班额更大的学校，效果会打折。承认这一点，比宣称可推广更诚实。

5.6 研究局限

局限至少有五条。单校、两班、非随机分班，代表性有限；同一位教师任教，霍桑效应与新异效应难以排除，虽然对照班也录像、也接受同等课时，教师投入的差异仍无法完全对等；周期只有 16 周，技能之外的迁移看不出来；识别只在五个拳术动作上验证，器械、对练、对抗都没覆盖；成本只算硬件，没算教师时间，后者更贵。

设计与测量上还有几处欠账。干预是一个打包方案，诊断、分组、任务卡、抽检绑在一起，没有设“只分组、不诊断”的第三组，技能提升里多少属于技术、多少属于组织方式，本研究分不开。达标线定在 80 分，换一条线率差会变，敏感性分析没有做。学生那一端只有量表与成绩，没有访谈，他们对被拍摄、被提示的感受，这里说不清。

还有一条伦理顾虑。系统记录学生动作，即便面部模糊、数据校内存储，学生仍会感到被注视。知情同意写明可随时退出，实际退出 2 人，数据未纳入分析。比例不高，但足以提醒：技术进课堂，先解决“愿不愿意被看”，再谈“看得准不准”。

6 结论与建议

第一，动作识别在武术教学里的合理定位是“形的测量仪”。它能给关节角、节拍、重心这些可比较的量，给不出劲力、意图与对抗时机。把定位说清楚，才能避免两种误用：当自动裁判用，或嫌不准弃之不用。

第二，16 周准实验显示，这套路径对技能成绩、达标率、练习密度、课外练习时长与技能自我效能均有正向作用，效应量 d=0.60~1.87；对理论测验没有作用。技术的作用范围是动作，不是知识，这一点应写进方案说明，免得学校期望错位。

第三，效果的关键变量是处方与撤除，不是精度。90.2% 够用，因为教学要方向不要读数；决定成败的是偏差能否聚成三类可执行处方、反馈能否按计划减量、教师有没有否决权。这三件事都不在算法里。

据此提五条建议：先定教学问题再选技术，要解决“看不见”就别买评分系统；把阈值制定权交给教研组，让规则停在教师能解释的层面；反馈设计从第一天就写撤除计划；教师培训的重心放在“什么时候不该信系统”；把阈值表与处方模板做成可共享的公共件。一所学校从零定阈值要花一个学期，十所学校各定一遍是重复劳动，由区域教研共同体出一版底表、各校按学情微调，门槛能降下来。

还有一句方法上的话。教育技术研究容易陷入“展示—验证”的循环：先展示技术能做什么，再找一批学生证明有效。这种研究不缺，缺的是把技术放回真实约束里，看它在班额、课时、教师精力的挤压下还剩多少用。本研究试着做后者，所以报告了不少不好看的数据：误报率 5.7%、8 人撤除后回落、理论成绩毫无变化。

参 考 文 献

[1] 董倩,刘阳.教育数字化赋能学校体育发展:历史、现状与未来[J].体育科学,2024,44(8):50-60.

[2] 于素梅,王晓燕.教育数字化背景下学校体育本质再认识[J].上海体育大学学报,2025,49(8):1-10.

[3] 张鑫淼,朱青,蔡玉军.人工智能时代的体育学习变革:典型样态、现实挑战与突破路径[J].上海体育大

学学报,2025,49(12):37-48.

[4] 郭江浩,熊文.人工智能赋能学校体育的限度及其超越:基于技术现象学的考察[J].体育学刊,2025,32(2):87-96.

[5] 曾天雪,王飞,王春昊.算法社会武术技术研究的科学化:长兵“攻击前手”技术数据分析及其启示[J].武汉体育学院学报,2023,57(9):54-61.

[6] 刘文武,王梦东.提升武术教学趣味性:问题审视、致因分析、改革对策[J].体育学刊,2024,31(5):109-114.

[7] PANG Y Q, ZHANG K Q, LI F M. Explainable quality assessment of effective aligned skeletal representations for martial arts movements by multi-machine learning decisions[J]. Scientific Reports, 2025, 15(1): 323.

[8] SIANTO L P, KRISTIAN Y. Automated performance analysis of Wushu movements using pose estimation and dynamic time warping[C]//2025 International Seminar on Intelligent Technology and Its Applications (ISITIA). Piscataway: IEEE, 2025: 783-788.

[9] YUAN L, HE Z, WANG Q, et al. Improving small-scale human action recognition performance using a 3D heatmap volume[J]. Sensors, 2023, 23(14): 6364.

[10] HAMILTON R I, GLAVCHEVA-LALEVA Z, HAQUE MILON M I, et al. Comparison of computational pose estimation models for joint angles with 3D motion capture[J]. Journal of Bodywork and Movement Therapies, 2024, 40: 315-319.

[11] WEAKLEY J, COWLEY N, SCHOENFELD B J, et al. The effect of feedback on resistance training performance and adaptations: a systematic review and meta-analysis[J]. Sports Medicine, 2023, 53(9): 1789-1803.

[12] 董新有,杜宴林.体育智能裁判系统对竞技体育的赋能、挑战与应对[J].武汉体育学院学报,2024,58(3):48-55.

[13] WULF G, SHEA C, LEWTHWAITE R. Motor skill learning and performance: a review of influential factors[J]. Medical Education, 2010, 44(1): 75-84.

Copyright © 2026 by author(s) and Global Academic Press Co., Limited.

This work is licensed under the Creative Commons Attribution-Non Commercial-Share Alike 4.0 International License. https://creativecommons.org/licenses/by-nc-sa/4.0/

